MASTER EN SCIENCES MATHÉMATIQUES

Méthodes intérieures du point proximal et du gradient pour l'optimisation convexe et conique

Lambert, Delphine

Award date:
2007

FUNDP
NAMUR

Faculté des Sciences
Département de Mathématique

Rempart de la Vierge, 8
B - 5000 Namur (Belgique)

Méthodes intérieures du point proximal et du gradient pour l'optimisation convexe et conique

Mémoire présenté pour l'obtention
du grade de
Licencié en Sciences Mathématiques
par

LAMBERT **Delphine**

Promoteur : STRODIOT Jean-Jacques

Année Académique 2006-2007

Résumé

Les méthodes intérieures du point proximal et du gradient (sous-gradient) ont déjà été beaucoup étudiées en optimisation convexe. Elles sont basées sur une mesure de proximité associée à la norme Euclidienne.

Dans ce travail, nous considérons à nouveau ces méthodes et nous introduisons une autre mesure de proximité qui nous permet d'éliminer les contraintes et de présenter des résultats de convergence globale similaires à ceux obtenus dans le cas sans contraintes.

Les résultats sont illustrés par des applications et exemples ainsi que par de nouveaux algorithmes simples pour les problèmes d'optimisation conique.

En particulier, nous trouvons une classe d'algorithmes du gradient intérieur qui fournit un taux de convergence global estimé de l'ordre de k^{-2}.

Mots-clé : optimisation convexe, algorithmes du gradient/sous-gradient intérieur, distances proximales, optimisation conique et convergence.

Abstract

Gradient (subgradient) and proximal interior methods for convex minimization have already been much studied. They are based on a proximity measure associated with the Euclidian norm.

In this work, we consider these methods again and we introduce another proximity measure which allows us to eliminate the constraints and to present global convergence results similar to those in the unconstrained case.

The results are illustrated with applications and examples, including some new simple algorithms for conic optimization problems.

In particular, we derive a class of interior gradient algorithms which exhibits an $\mathcal{O}(k^{-2})$ global convergence rate estimate.

Keywords : convex optimization, interior gradient/subgradient algorithms, proximal distance, conic optimization and convergence.

A la fin de ces années d'études, laborieuses par moments, mais passionnantes et amusantes à d'autres, je tiens à remercier toutes les personnes qui m'ont soutenue...

Tout d'abord, concernant ce mémoire qui représente l'accomplissement de mes études, j'adresse un très grand merci à mon promoteur, Jean-Jacques Strodiot, pour l'encadrement et le temps qu'il m'a consacré tout au long de l'année pour mener à bien ce projet.

Ensuite, je tiens à remercier tous les assistants et, en particulier, Caroline et Sebastian pour leur disponibilité, leur soutien et leur aide au fil des années pour résoudre les petits problèmes.

Je remercie aussi mes parents et Sébastien pour leur écoute, leur patience, leur soutien et surtout leurs encouragements dans les moments plus difficiles.

Table des matières

Notations

S	sous-ensemble non vide de $\mathbb{R}^n$
$\overline{S}$	fermeture de S
A^T	transposée de la matrice A
(x^n)	suite
(α_n)	suite de nombres
$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$	fonction convexe et semi-continue inférieurement
$f \in C^1$	fonction une fois continûment différentiable
$f \in C^2$	fonction deux fois continûment différentiable
dom f	domaine de la fonction f
$\nabla f(x)$	gradient de f en x
$\nabla^2 f(x)$	matrice hessienne de f en x
$\partial f(x)$	sous-différentiel de f au point x
$\partial_\varepsilon f(x)$	sous-différentiel approximé
	ou ε-sous-différentiel de f au point x
δ_S	fonction indicatrice de S
ri S	intérieur relatif de S
co S	enveloppe convexe de S
aff S	enveloppe affine de S

$N_S(x)$	cône normal de S en $x \in S$
$\mathbb{R}^n_+$	ensemble des n-vecteurs à composantes positives
$\mathbb{R}^n_{++}$	ensemble des n-vecteurs à composantes strictement positives
$\Gamma_0(\mathbb{R}^n)$	ensemble des fonctions convexes, propres et fermées sur $\mathbb{R}^n$
$\mathcal{V}$	ensemble des points x tels que $Ax = b$
$\mathcal{V}_0$	ensemble des points x tels que $Ax = 0$
d	distance proximale
H	distance proximale induite
D_h	distance proximale de Bregman

Introduction

Dans ce travail, nous considérons le problème de minimisation convexe suivant :

$$f_* = \inf\{f(x) \mid x \in \overline{C}\}, \qquad (P)$$

où $\overline{C}$ représente la fermeture de C, un ensemble convexe ouvert non vide de $\mathbb{R}^n$ et où $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ est une fonction convexe, propre et semi-continue inférieurement.

Pour résoudre le problème (P), nous étudions deux schémas itératifs relativement proches.

Le premier est basé sur la méthode proximale. Etant donné une mesure de proximité d, il consiste à générer une suite (x^k) via l'itération

$$x^k \in \arg\min\{\lambda_k f(x) + d(x, x^{k-1}) \mid x \in \overline{C}\}, \quad k = 1, 2, \ldots \quad (\lambda_k > 0).$$

Le second est basé sur la méthode du sous-gradient et produit une suite (x^k) via

$$x^k \in \arg\min\{\lambda_k \langle g^{k-1}, x\rangle + d(x, x^{k-1}) \mid x \in \overline{C}\}, \quad k = 1, 2, \ldots$$

où $\langle \cdot, \cdot \rangle$ est un produit intérieur sur $\mathbb{R}^n$ et g^{k-1} le sous-gradient de la fonction f au point x^{k-1}.

Nous analysons plus particulièrement le taux de convergence des deux méthodes et nous proposons un schéma qui améliore leur efficacité.

Ce mémoire est basé principalement sur l'article de Alfred Auslender et Marc Teboulle [2].

Chapitre 1

Préliminaires

1.1 Notions d'analyse

1.1.1 Suites de nombres réels

Notions de convergence

Définition 1.1.1 *Une suite* (x^n) *converge vers un point* x *, i.e.* $x^n \to x$

$\Leftrightarrow \forall \varepsilon > 0, \exists\ n_\varepsilon \in I\!N$ *tel que* $\forall n \geq n_\varepsilon, x^n \in]x-\varepsilon, x+\varepsilon[$ *i.e.* $|x^n - x| < \varepsilon$.

Théorème 1.1.1 ***Théorème de l'étau***

Si (a^n) *et* (b^n) *convergent vers* x *et si* $\exists\ n_0 \in I\!N$ *tel que* $\forall n \geq n_0,\quad a^n \leq x^n \leq b^n,\quad$ *alors* $(x^n) \to x$.

Valeurs d'adhérence et sous-suites

Définition 1.1.2 (y^n) *est une sous-suite de* $(x^n) \Leftrightarrow y^n = x^{q_n},\ \forall n \in I\!N$ *où* $(q_n)_{n \in I\!N}$ *est une suite strictement croissante de naturels.*

Définition 1.1.3 *x est une valeur d'adhérence de (x^n)*
$\Leftrightarrow$ x est limite d'une sous-suite de (x^n).

Les propriétés suivantes sont aussi intéressantes.

Proposition 1.1.1 *Si (x^n) est une suite qui converge vers x,*

alors, x est l'unique valeur d'adhérence de (x^n).

Théorème 1.1.2 ***Théorème de Bolzano-Weierstrass***

Soit (x^n) une suite bornée.

Alors, (x^n) admet une sous-suite convergente.

Rappelons également l'inégalité de Cauchy-Schwarz :

$$|\langle x, y\rangle| \leq \|x\|\|y\|.$$

1.1.2 Dérivées

Théorème 1.1.3 ***Théorème de Rolle***

Soit $f : [a,b] \to \mathbb{R}$ continue sur $[a,b]$, dérivable sur $[a,b]$ et telle que $f(a) = f(b)$.

Alors, $\exists\, c \in]a,b[: f'(c) = 0$.

Théorème 1.1.4 ***Théorème des accroissements finis***

Soit $f : [a, b] \to \mathbb{R}$ *continue sur* $[a, b]$ *et dérivable sur* $]a, b[$.

Alors, $\exists\, c \in]a, b[: f'(c) = \dfrac{f(b) - f(a)}{b - a}$.

1.2 Notions de topologie

Définition 1.2.1 *Soit* X *un ensemble non vide.*
Une classe $\mathcal{T} \subset \mathcal{P}(X)$ *est une topologie sur* X *si et seulement si*

(O_1) $\quad \emptyset$ *et* $X \in \mathcal{T}$;

(O_2) $\quad \forall (G_i)_{i \in I} \subset \mathcal{T} : \bigcup_{i \in I} G_i \in \mathcal{T}$;

(O_3) $\quad$ *Si* G_1 *et* $G_2 \in \mathcal{T}$, *alors* $G_1 \cap G_2 \in \mathcal{T}$.

Nous disons alors que $(X, \mathcal{T})$ *est un espace topologique et les éléments de* $\mathcal{T}$ *sont appelés des ouverts notés par* G .

Définition 1.2.2 *Soit* $(X, \mathcal{T})$ *un espace topologique.*
$F \subset X$ *est fermé si et seulement si* $F^c \in \mathcal{T}$ *où* F^c *est l'ensemble complémentaire de* F .
L'ensemble des fermés est noté $\mathcal{F}$ *et un fermé est noté* F .

Définition 1.2.3 *La fermeture d'un ensemble* A *est notée* $\overline{A}$ *et est définie comme suit :*

$$\overline{A} := \cap \{F \in \mathcal{F} : F \supset A\}.$$

Un point p *appartenant à* $\overline{A}$ *est appelé point de fermeture.*

Définition 1.2.4 *Un ensemble X est dit compact si et seulement si tout recouvrement par des ouverts de X admet un sous-recouvrement fini.*

1.3 Fonctions convexes

1.3.1 Définitions et propriétés

Définition 1.3.1

1. *Soit C un sous-ensemble non vide de $\mathbb{R}^n$. Alors, C est convexe si*

$$\forall x, y \in C, \forall \alpha \in]0,1[, \quad \alpha x + (1-\alpha)y \in C.$$

2. *Soit C un sous-ensemble convexe non vide de $\mathbb{R}^n$.*
Une fonction $f : C \to \mathbb{R}$ est convexe sur C si

$$\forall x, y \in C, \forall \alpha \in]0,1[, \quad f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y).$$

3. *Soit C un sous-ensemble convexe non vide de $\mathbb{R}^n$.*
Une fonction $f : C \to \mathbb{R}$ est strictement convexe sur C si

$$\forall x, y \in C, x \neq y, \forall \alpha \in]0,1[, \quad f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y).$$

4. *Soit C un sous-ensemble convexe non vide de $\mathbb{R}^n$.*
Une fonction $f : C \to \mathbb{R}$ est fortement convexe de module $m > 0$ si $f - \frac{m}{2}\|.\|^2$ est convexe sur C.

5. *f est concave (resp. strictement concave, fortement concave) sur C si $-f$ est convexe (resp. strictement convexe, fortement convexe).*

Les propriétés des fonctions convexes suivantes sont intéressantes pour la suite.

Proposition 1.3.1 *Soit C un sous-ensemble convexe non vide de $\mathbb{R}^n$.*

Soit $f : C \to \mathbb{R}$ avec f de classe C^1 sur C. Alors,

$$f \text{ est convexe sur } C \Leftrightarrow \forall x, y \in C,\ f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

où $\nabla f(x)$ représente le gradient de f en x.

Proposition 1.3.2 *Les assertions suivantes sont équivalentes :*

1. *f est fortement convexe sur C de module $m > 0$*
2. $\forall x, y \in C,\ f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{m}{2}\|y - x\|^2$
3. $\forall x \in C,\ \forall d \in \mathbb{R}^n,\ d^T \nabla^2 f(x) d > m\|d\|^2$

Théorème 1.3.1 *Soit $f : C \subset \mathbb{R}^n \to \mathbb{R}$ une fonction fortement convexe.*

Alors, f admet un minimum unique sur C.

Définition 1.3.2 *Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.*
Le domaine de f est l'ensemble

$$dom\ f = \{x \in \mathbb{R}^n \mid f(x) < +\infty\}.$$

La fonction f est propre si dom f est non vide.

Définition 1.3.3 *Soit C un sous-ensemble convexe non vide de $\mathbb{R}^n$.*
Nous définissons alors l'enveloppe convexe de C comme étant le plus petit sous-espace convexe contenant C.
L'enveloppe convexe de C est notée co C.

Nous pouvons définir de la même manière *l'enveloppe affine* de C que nous noterons *aff* C.

Définition 1.3.4 *Une fonction* $f : I\!R^n \to I\!R \cup \{+\infty\}$ *est dite coercive sur un sous-espace fermé non vide* C *de* $I\!R^n$ *si*

$$\lim_{\substack{\|x\| \to +\infty \\ x \in C}} f(x) = +\infty.$$

Définition 1.3.5 *L'intérieur relatif d'un sous-espace convexe* C *de* $I\!R^n$ *est l'intérieur de* C *pour la topologie relative à l'enveloppe affine de* C, *i.e.*

$$x \in \ ri\, C \Leftrightarrow \begin{cases} x \in \ aff\, C \\ \exists\ \delta > 0 \ \textit{tel que} \ aff\, C \cap B(x, \delta) \subseteq C \end{cases}$$

1.3.2 Fonctions convexes spéciales

Voici des classes de fonctions que nous rencontrerons plus tard : les fonctions affines et les fonctions semi-continues inférieurement.

Définition 1.3.6 *Soient* $a \in I\!R^n$ *et* $b \in I\!R$.
La fonction $f : I\!R^n \to I\!R$ *définie par* $f(x) = \langle a, x \rangle - b \quad \forall x \in I\!R^n$ *est appelée fonction affine sur* $I\!R^n$.
Si $b = 0$, *alors* $f(x) = \langle a, x \rangle$ *est dite linéaire.*

Définition 1.3.7 *Soit* $f : I\!R^n \to I\!R \cup \{+\infty\}$ *une fonction propre.*
f *est dite semi-continue inférieurement (sci) en* $x \in I\!R^n$ *si*

$$\liminf_{y \to x} f(y) \geq f(x).$$

1.4 Sous-différentiabilité

Soient $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ et $x \in$ int (dom f).
Le but du calcul différentiel est d'approximer f au voisinage de x par une fonction linéaire.

Définition 1.4.1 *La fonction f est différentiable en x s'il existe $l_x : \mathbb{R}^n \to \mathbb{R}$ linéaire telle que*

$$f(x+h) = f(x) + l_x(h) + (o\|h\|).$$

Ici, $g(h) = o(\|h\|)$ quand $h \to 0$ signifie que $\lim_{h\to 0} \frac{g(h)}{\|h\|} = 0$.

De plus, dans ce cas, l_x est unique et s'exprime par

$$l_x(h) = \langle \nabla f(x), h \rangle = \sigma_{\{\nabla f(x)\}}(h).$$

Le vecteur $\nabla f(x)$ est appelé gradient de f en x.

Définition 1.4.2 *Soient $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $x_0 \in$ dom f et $d \in \mathbb{R}^n$. La dérivée directionnelle de f en x_0 dans la direction d est*

$$f'(x,d) = \lim_{t \searrow 0} \frac{f(x_0 + td) - f(x_0)}{t}$$

si la limite existe ($-\infty$ et $+\infty$ étant autorisés).

Définition 1.4.3 *Soit C un sous-ensemble non vide de $\mathbb{R}^n$. Le fonction support de C est définie par*

$$\sigma_C : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \qquad \sigma_C(x) = \sup_{s \in C} \langle x, s \rangle.$$

Définition 1.4.4 *Le sous-différentiel $\partial f(x)$ de f en x est l'ensemble compact convexe non vide de $\mathbb{R}^n$ pour lequel la fonction support est $f'(x,.)$, i.e.*

$$\partial f(x) = \{s \in \mathbb{R}^n \mid \langle s, d \rangle \leq f'(x,d) \quad \forall d \in \mathbb{R}^n\}.$$

Tout vecteur $s \in \partial f(x)$ est appelé sous-gradient de f en x.

En d'autres mots, nous avons l'égalité importante

$$f'(x,d) = \sup_{s \in \partial f(x)} \langle s, d \rangle.$$

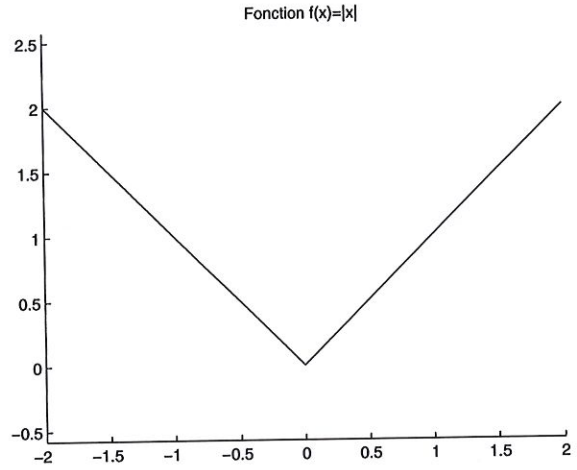
Voici une propriété que nous utiliserons régulièrement :

$$s \in \partial f(x_0) \quad \Leftrightarrow \quad \forall x \in \mathbb{R}^n, f(x) \geq f(x_0) + \langle s, x - x_0 \rangle.$$

Nous pouvons en donner l'interprétation géométrique suivante : l'inégalité définissant le $s \in \partial f(x_0)$ signifie que s est la pente d'une fonction affine qui minimise f et qui passe par le point $(x_0, f(x_0))$.

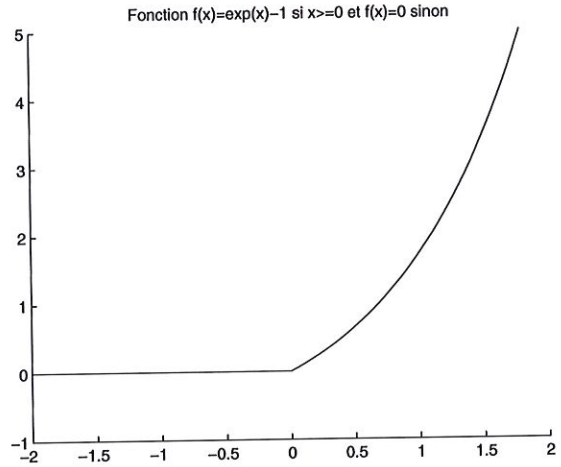
Exemple 1 *Pour illustrer cette interprétation, nous considérons la fonction* $f(x) = |x|$. *Alors, nous remarquons facilement sur le graphique de cette fonction que*

$$\partial f(0) = [-1, 1], \quad \partial f(x_0) = \{1\} \text{ si } x_0 > 0 \text{ et } \partial f(x_0) = \{-1\} \text{ si } x_0 < 0.$$



Exemple 2 *Pour la fonction* $f(x) = e^x - 1$ *si* $x \geq 0$ *et* 0 *si* $x < 0$, *nous 'obtenons*

$$\partial f(0) = [0, 1], \quad \partial f(x_0) = \{e^{x_0}\} \text{ si } x_0 > 0 \text{ et } \partial f(x_0) = \{0\} \text{ si } x_0 < 0.$$



1.5 Sous-différentiel approximé

Définition 1.5.1 *Soit* $f \in \Gamma_0(\mathbb{R}^n)$ *où* $\Gamma_0(\mathbb{R}^n)$ *est l'ensemble des fonctions convexes propres fermées sur* $\mathbb{R}^n$.
Soient $\varepsilon \geq 0$ *et* $x \in$ *dom* f .
Un vecteur $s \in \mathbb{R}^n$ *est appelé un* ε *-sous-gradient (ou sous-gradient approximé) de* f *en* x *si*

$$\forall y \in \text{dom } f, \qquad f(y) \geq f(x) + \langle s, y - x \rangle - \varepsilon.$$

L'ensemble des ε *-sous-gradients de* f *en* x *est noté* $\partial_\varepsilon f(x)$ *et est appelé le* ε *-sous-différentiel de* f *en* x .
Par convention, $\partial_\varepsilon f(x) = \emptyset$ *quand* $x \notin$ *dom* f .

Chapitre 2

Cadre général pour les méthodes proximales intérieures

2.1 Introduction

Soient C un ensemble ouvert convexe non vide de $\mathbb{R}^n$ et $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe, propre et semi-continue inférieurement.

Considérons le problème

$$f_* = \inf\{f(x) \mid x \in \overline{C}\}. \qquad (P)$$

Nous faisons, à travers tout ce travail, les hypothèses suivantes sur (P) :

1. dom $f \cap C \neq \emptyset$
2. $-\infty < f_*$

Pour commencer, nous allons étudier le comportement du schéma itératif proximal suivant pour résoudre (P) :

$$x^k \in \arg\min\{\lambda_k f(x) + d(x, x^{k-1}) \mid x \in \overline{C}\}, \quad k = 1, 2, \ldots \qquad (\lambda_k > 0),$$

où d est une certaine distance proximale.

L'objectif de ce travail est de développer un schéma général pour analyser la convergence des méthodes dans des cadres variés.

Etant donné le problème d'optimisation (P), les étapes nécessaires pour atteindre ce but sont :

- le choix d'une distance proximale d appropriée permettant d'éliminer les contraintes ;
- étant donné d, la recherche d'une distance proximale induite H qui contrôlera le comportement de la méthode résultante.

Commençons par définir une distance proximale d appropriée pour le problème (P).

Définition 2.1.1 *Soit C un ouvert convexe non vide de $\mathbb{R}^n$.*
Une fonction $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ est appelée distance proximale relative à C si, pour chaque $y \in C$, elle satisfait aux propriétés suivantes :

(P_1) $d(\cdot, y)$ est convexe, propre, sci et C^1 sur C ;

(P_2) dom $d(\cdot, y) \subset \overline{C}$ et dom $\partial_1 d(\cdot, y) = C$, où $\partial_1 d(\cdot, y)$ représente le sous-différentiel de la fonction $d(\cdot, y)$ par rapport à la première variable ;

(P_3) $d(\cdot, y)$ est coercive sur $\mathbb{R}^n$;

(P_4) $d(y, y) = 0$.

Nous notons $\mathcal{D}(C)$ la famille des fonctions d satisfaisant aux propriétés de cette définition.

La propriété (P_1) est nécessaire pour préserver la convexité de $d(\cdot, y)$, (P_2) force l'itéré x^k à rester dans C et (P_3) est utilisée pour garantir l'existence d'un tel itéré.
Pour chaque $y \in C$, notons $\nabla_1 d(\cdot, y)$ le gradient de la fonction $d(\cdot, y)$ par rapport à la première variable.
Notons aussi que, par définition, $d(\cdot, \cdot) \geq 0$ et, par (P_4), que le minimum global de $d(\cdot, y)$ est obtenu en y, ce qui montre que $\nabla_1 d(y, y) = 0$.

Lemme 2.1.1 *Soit* $x \in \overline{C}$ *et* $x' \in ri\ C$.

Alors, la demi-droite

$$]x, x'] = \{\alpha x + (1-\alpha)x' \mid 0 \leq \alpha < 1\}$$

est contenue dans $ri\ C$.

Preuve

Prenons $x'' = \alpha x + (1-\alpha)x'$ avec $1 > \alpha \geq 0$.
Supposons, sans perte de généralité, que aff $C = \mathbb{R}^n$.
Comme $x \in \overline{C}$, pour tout $\varepsilon > 0, x \in C + B(0,\varepsilon)$ et nous pouvons écrire

$$\begin{aligned} B(x'',\varepsilon) &= \alpha x + (1-\alpha)x' + B(0,\varepsilon) \\ &\subset \alpha C + (1-\alpha)x' + (1+\alpha)B(0,\varepsilon) \\ &= \alpha C + (1-\alpha)\left\{x' + B\left(0, \frac{1+\alpha}{1-\alpha}\varepsilon\right)\right\}. \end{aligned}$$

Comme $x' \in$ int C, choisir ε assez petit implique que $x' + B\left(0, \frac{1+\alpha}{1-\alpha}\varepsilon\right) \subset C$. Alors, nous avons

$$B(x'',\varepsilon) \subset \alpha C + (1-\alpha)C = C. \qquad \blacksquare$$

Proposition 2.1.1 *Soient* $f_1, \ldots, f_m$ *des fonctions convexes propres sur* $\mathbb{R}^n$.

Soit $f = f_1 + \ldots + f_m$. *Alors,*

$$\partial f(x) \supset \partial f_1(x) + \ldots + \partial f_m(x), \qquad \forall x.$$

Si les ensembles convexes $ri\ (\ dom\ f_i)$, $i = 1, \ldots, m$ *ont un point commun, alors,*

$$\partial f(x) = \partial f_1(x) + \ldots + \partial f_m(x), \qquad \forall x.$$

Cette condition pour l'égalité peut être simplifiée si certaines fonctions, à savoir $f_1, \ldots, f_k$, *sont polyédriques : alors, il suffit que les ensembles* $dom\ f_i$, $i = 1, \ldots, k$ *et* $ri\ (\ dom\ f_i)$, $i = k+1, \ldots, m$ *aient un point en commun.*

Proposition 2.1.2 *Soit* $d \in \mathcal{D}(C)$ *et, pour tout* $y \in C$, *considérons le problème d'optimisation*

$$f_*(y) = inf\{f(u) + d(u, y) \mid u \in \mathbb{R}^n\}. \qquad (P(y))$$

Alors, l'ensemble optimal $S(y)$ *de* $P(y)$ *est non vide et compact, et, pour chaque* $\varepsilon \geq 0$, *il existe* $u(y) \in C$, $g \in \partial_\varepsilon f(u(y))$ *tels que*

$$g + \nabla_1 d(u(y), y) = 0 \qquad (2.1)$$

où $\partial_\varepsilon f(u(y))$ *représente le sous-différentiel approximé de* f *en* $u(y)$.

Pour un tel $u(y) \in C$, *nous avons*

$$f(u(y)) + d(u(y), y) \leq f_*(y) + \varepsilon. \qquad (2.2)$$

Preuve

Posons $t(u) = f(u) + d(u,y) + \delta_{\overline{C}}(u)$ où $\delta_{\overline{C}}(u)$, la fonction indicatrice, est définie par

$$\delta_{\overline{C}}(u) = \begin{cases} 0 & \text{si} \quad u \in \overline{C}; \\ +\infty & \text{sinon.} \end{cases}$$

Alors, par (P_2), nous avons $f_*(y) = \inf \{t(u) \mid u \in \mathbb{R}^n\}$.

De plus, comme f_* est fini, il suit de (P_3) que $t(\cdot)$ est coercive.

Donc, comme $t(\cdot)$ est une fonction convexe, propre et sci, il suit que $S(y)$ est non vide et compact. En effet, $S(y)$ est non vide car dom t est non vide ($t(\cdot)$ est propre).

Par les conditions d'optimalité, pour chaque $u(y) \in S(y)$, nous avons $0 \in \partial t(u(y))$.

Comme dom $f \cap C \neq \emptyset$ et comme C est ouvert, nous obtenons, par la proposition 2.1.1,

$$\partial t(u) = \partial f(u) + \nabla_1 d(u,y) + N_{\overline{C}}(u) \quad \forall u.$$

En effet, comme $t(u) = f(u) + d(u,y) + \delta_{\overline{C}}(u)$, si nous regardons le domaine, nous avons dom $t(u) =$ dom $f \cap C$.
En prenant l'intérieur relatif, nous avons ri $t(u) =$ ri dom $f \cap C$.
Or, par le lemme 2.1.1, il est possible de montrer que dom $f \cap C \neq \emptyset \Rightarrow$ ri dom $f \cap C \neq \emptyset$.
Donc, il existe bien un point dans l'intersection des intérieurs relatifs et la proposition 2.1.1 est applicable.

Comme dom $\partial_1 d(\cdot, y) = C$, il suit que $u(y) \in C$ et donc $N_{\overline{C}}(u(y)) = \{0\}$.

Dès lors, la propriété (2.1) a lieu pour $\varepsilon = 0$ avec $g \in \partial f(u(y))$.
Pour $\varepsilon > 0$, la propriété (2.1) a lieu pour une paire $(u(y), g)$ car $\partial f(u(y)) \subset \partial_\varepsilon f(u(y))$ et donc, la première partie de la proposition est prouvée.

Finalement, comme pour chaque $y \in C$, la fonction $d(\cdot, y)$ est convexe et comme $g \in \partial_\varepsilon f(u(y))$, nous avons

$$f(u) + d(u, y) \geq f(u(y)) + d(u(y), y) + \langle g + \nabla_1 d(u(y), y), u - u(y) \rangle - \varepsilon$$

de telle sorte que

$$\begin{aligned} f_*(y) &= \inf\{f(u) + d(u, y) \mid u \in \overline{C}\} \\ &\geq f(u(y)) + d(u(y), y) - \varepsilon \end{aligned}$$

et l'inégalité (2.2) est bien obtenue. ∎

Grâce à cette proposition, l'algorithme de base, présenté dans la section suivante, est bien défini.

2.2 Algorithme proximal intérieur

Algorithme Proximal Intérieur (IPA)

Etant donné $d \in \mathcal{D}(C)$, commencer avec un point $x^0 \in C$ et, pour $k = 1, 2, \ldots,$ avec $\lambda_k > 0$, $\varepsilon_k \geq 0$, générer une suite

$$(x^k) \in C \text{ avec } g^k \in \partial_{\varepsilon_k} f(x^k) \tag{2.3}$$

telle que

$$\lambda_k g^k + \nabla_1 d(x^k, x^{k-1}) = 0. \tag{2.4}$$

L'algorithme IPA peut être vu comme une méthode proximale intérieure approximée quand ε_k est strictement positif pour tout $k \in \mathbb{N}$. Cette méthode devient exacte quand ε_k égale zéro pour tout $k \in \mathbb{N}$.

L'étape suivante est d'associer à chaque distance $d \in \mathcal{D}(C)$ donnée une distance proximale correspondante satisfaisant certaines propriétés nécessaires à l'analyse de la convergence de l'algorithme IPA.

Définition 2.2.1 *Soit C un ouvert convexe non vide de $\mathbb{R}^n$ et soit une distance proximale $d \in \mathcal{D}(C)$.*
Une fonction $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ est appelée distance proximale induite à d si

1. *H est à valeurs finies sur $C \times C$;*
2. *pour tout $a \in C$,*
$$H(a, a) = 0 \ ; \tag{2.5}$$
3. *pour tout $a, b \in C$,*
$$\langle c - b, \nabla_1 d(b, a) \rangle \leq H(c, a) - H(c, b) \quad \forall c \in C. \tag{2.6}$$

Nous écrivons $(d, H) \in \mathcal{F}(C)$ pour quantifier le triplet $[C, d, H]$ qui satisfait aux propriétés de la définition 2.2.1.

Définition 2.2.2 *Soit C un ouvert convexe non vide de $\mathbb{R}^n$ et soit une distance proximale $d \in \mathcal{D}(C)$.*
Une fonction $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ est appelée distance proximale induite à d si

1. *H est à valeurs finies sur $\overline{C} \times C$;*
2. *pour tout $a \in C$,*
$$H(a,a) = 0 \ ;$$
3. *pour tout $a, b \in C$,*
$$\langle c - b, \nabla_1 d(b,a)\rangle \leq H(c,a) - H(c,b) \quad \forall c \in \overline{C} \ ;$$
4. *$H(c,\cdot)$ est coercive sur C.*

Nous écrirons $(d,H) \in \mathcal{F}(\overline{C})$ pour quantifier le triplet $[\overline{C}, d, H]$ qui satisfait aux propriétés de cette définition.

Clairement, nous avons $\mathcal{F}(\overline{C}) \subset \mathcal{F}(C)$.

Notons que l'algorithme proximal classique PA correspond au cas spécial où $C = \overline{C} = \mathbb{R}^n$, $d(x,y) = \frac{1}{2}\|x - y\|^2$ et avec la distance proximale induite H étant exactement d. Cet algorithme satisfait (2.6) grâce à l'égalité bien connue

$$\|z - x\|^2 = \|z - y\|^2 + \|y - x\|^2 + 2\langle z - y, y - x\rangle.$$

En effet, nous avons

$$\begin{aligned}
\langle c - b, \nabla_1 d(b,a)\rangle &\leq \frac{1}{2}\|c - a\|^2 - \frac{1}{2}\|c - b\|^2 && \forall c \in C \\
&= \frac{1}{2}\|c - b\|^2 + \frac{1}{2}\|b - a\|^2 + 2.\frac{1}{2}\langle c - b, b - a\rangle \\
&\quad - \frac{1}{2}\|c - b\|^2 && \forall c \in C \\
&= \frac{1}{2}\|b - a\|^2 + \langle c - b, b - a\rangle.
\end{aligned}$$

Or, puisque $\nabla_1 d(b,a) = b - a$, nous obtenons bien le résultat souhaité

$$\langle c - b, \nabla_1 d(b,a)\rangle \leq \frac{1}{2}\|c - a\|^2 - \frac{1}{2}\|c - b\|^2 \qquad \forall c \in C.$$

L'algorithme IPA avec $d \equiv H$ sera appelé algorithme *self-proximal.*

Plusieurs exemples de méthodes self-proximales seront donnés en fin de chapitre.

Comme nous pourrons aussi le voir plus loin, les propriétés de la fonction H associée à d émergent naturellement de l'analyse de l'algorithme classique PA comme donné dans [14] et sont étendues à des classes spécifiques de l'algorithme IPA dans [3],[10],[29].
En se basant sur ces travaux, nous pouvons facilement obtenir des taux de convergence estimés globaux ainsi que la convergence aux valeurs d'adhérence de la suite produite par l'algorithme IPA.
Pour transformer le taux de convergence global de la suite (x^k) en une solution optimale de (P), des hypothèses supplémentaires sur la distance proximale induite H, apparentées aux propriétés des normes, seront nécessaires.

Avant de donner nos résultats de convergence, rappelons les propriétés bien connues des suites positives qui seront utilisées tout au long de ce travail.

Lemme 2.2.1 *Soient* (v_k), (γ_k) *et* (β_k) *des suites positives de nombres réels satisfaisant*

$$v_{k+1} \leq (1+\gamma_k)v_k + \beta_k$$

et telles que $\sum_{k=1}^{\infty} \beta_k < \infty$ *et* $\sum_{k=1}^{\infty} \gamma_k < \infty$.

Alors, la suite (v_k) *converge.*

Lemme 2.2.2 *Soient* (λ_k) *une suite de nombres réels positifs,* (a_n) *une suite de nombres réels et* $b_n := \frac{1}{\sigma_n} \sum_{k=1}^{n} \lambda_k a_k$ *où* $\sigma_n = \sum_{k=1}^{n} \lambda_k$.

Si $\sigma_n \to \infty$,

1. $\liminf a_n \leq \liminf b_n \leq \limsup b_n \leq \limsup a_n$
2. $\lim b_n = a$ *chaque fois que* $\lim a_n = a$.

Lemme 2.2.3 *Soit f une fonction convexe propre et fermée.*

Si l'ensemble niveau $\{x \mid f(x) \leq \alpha\}$ est non vide et borné pour un certain α, alors, il est borné pour tout α.

Une preuve de ce lemme peut être trouvée dans [23]. ∎

Théorème 2.2.1 *Soit $(d, H) \in \mathcal{F}(C)$ et soit (x^k) une suite générée par l'algorithme IPA.*

Posons $\sigma_n = \sum_{k=1}^{n} \lambda_k$.

Alors, nous avons :

1. $f(x^n) - f(x) \leq \frac{1}{\sigma_n} H(x, x^0) + \frac{1}{\sigma_n} \sum_{k=1}^{n} \sigma_k \varepsilon_k \quad \forall x \in C$.

2. *Si* $\lim_{n\to\infty} \sigma_n = +\infty$ *et si* $\varepsilon_k \to 0$,
   alors, $\liminf_{n\to\infty} f(x^n) = f_*$ *et la suite* $(f(x^k))$ *converge vers* f_*
   chaque fois que $\sum_{k=1}^{\infty} \varepsilon_k < \infty$.

3. *De plus, supposons que l'ensemble optimal X_* du problème (P) soit non vide et considérons les cas suivants :*

   (a) X_* *est borné,*

   (b) $\sum_{k=1}^{\infty} \lambda_k \varepsilon_k < \infty$ *et* $(d, H) \in \mathcal{F}(\overline{C})$.

   Alors, sous les hypothèses (a) ou (b), la suite (x^k) est bornée avec toutes ses valeurs d'adhérence dans X_*.

Preuve

• Pas 1

Par (2.4) et comme $g^k \in \partial_{\varepsilon_k} f(x^k)$, nous avons

$$\lambda_k(f(x^k) - f(x)) \leq \langle x - x^k, \nabla_1 d(x^k, x^{k-1})\rangle + \lambda_k \varepsilon_k \quad \forall x \in C. \tag{2.7}$$

En effet, par définition du sous-différentiel approximé

$$s \in \partial_{\varepsilon_k} f(x^k) \quad \Leftrightarrow \quad \forall y \in \text{dom } f, \quad f(y) \geq f(x^k) + \langle s, y - x^k\rangle - \varepsilon_k,$$

nous avons donc

$$\begin{aligned}
f(y) &\geq f(x^k) + \langle s, y - x^k\rangle - \varepsilon_k \\
-\langle s, y - x^k\rangle + \varepsilon_k &\geq f(x^k) - f(y) \\
\langle x^k - y, s\rangle + \varepsilon_k &\geq f(x^k) - f(y) \\
\lambda_k\langle x^k - y, s\rangle + \lambda_k\varepsilon_k &\geq \lambda_k(f(x^k) - f(y)) \\
\langle x^k - y, \lambda_k s\rangle + \lambda_k\varepsilon_k &\geq \lambda_k(f(x^k) - f(y)) \\
\langle y - x^k, -\lambda_k s\rangle + \lambda_k\varepsilon_k &\geq \lambda_k(f(x^k) - f(y)) \qquad \forall y \in \text{ dom } f.
\end{aligned}$$

Nous obtenons bien (2.7) car $\begin{cases} y = x\ ; \\ s = g^k \in \partial_{\varepsilon_k} f(x^k). \end{cases}$

En utilisant (2.6) aux points $c = x$, $a = x^{k-1}$, $b = x^k$, i.e.

$$\langle x - x^k, \nabla_1 d(x^k, x^{k-1})\rangle \leq H(x, x^{k-1}) - H(x, x^k) \quad \forall x \in C,$$

l'inégalité (2.7) implique que

$$\lambda_k(f(x^k) - f(x)) \leq H(x, x^{k-1}) - H(x, x^k) + \lambda_k\varepsilon_k \quad \forall x \in C. \tag{2.8}$$

En sommant sur $k = 1, \ldots, n$, nous obtenons

$$-\sigma_n f(x) + \sum_{k=1}^{n} \lambda_k f(x^k) \leq H(x, x^0) - H(x, x^n) + \sum_{k=1}^{n} \lambda_k\varepsilon_k. \tag{2.9}$$

En prenant $x = x^{k-1}$ dans (2.8), nous obtenons

$$f(x^k) - f(x^{k-1}) \leq \varepsilon_k. \tag{2.10}$$

En effet,

$$\begin{aligned}\lambda_k(f(x^k)-f(x^{k-1})) &\leq H(x^{k-1},x^{k-1})-H(x^{k-1},x^k)+\lambda_k\varepsilon_k\\ &\leq -H(x^{k-1},x^k)+\lambda_k\varepsilon_k\\ &\leq \lambda_k\varepsilon_k\end{aligned}$$

car $H(x,x)=0$ pour tout x et $H(\cdot,\cdot)\geq 0$.

En multipliant (2.10) par σ_{k-1} (avec $\sigma_0 \equiv 0$) et en sommant sur $k=1,\ldots,n$, i.e.

$$\sum_{k=1}^{n}\sigma_{k-1}f(x^k)-\sum_{k=1}^{n}\sigma_{k-1}f(x^{k-1})\leq\sum_{k=1}^{n}\sigma_{k-1}\varepsilon_k,$$

nous obtenons

$$\sigma_n f(x^n)-\sum_{k=1}^{n}\lambda_k f(x^k)\leq\sum_{k=1}^{n}\sigma_{k-1}\varepsilon_k. \tag{2.11}$$

En effet,

$$\begin{aligned}&\sum_{k=1}^{n}\sigma_{k-1}f(x^k)-\sum_{k=1}^{n}\sigma_{k-1}f(x^{k-1})\\ &\quad=\sum_{k=2}^{n}\sigma_{k-1}f(x^k)-\sum_{k=2}^{n}\sigma_{k-1}f(x^{k-1}) \qquad (\text{car } \sigma_0\equiv 0)\\ &\quad=\sigma_1f(x^2)-\sigma_1f(x^1)+\sigma_2f(x^3)-\sigma_2f(x^2)\\ &\qquad+\ldots+\sigma_{n-1}f(x^n)-\sigma_{n-1}f(x^{n-1})\\ &\quad=-\sigma_1f(x^1)+(\sigma_1-\sigma_2)f(x^2)+(\sigma_2-\sigma_3)f(x^3)\\ &\qquad+\ldots+(\sigma_{n-2}-\sigma_{n-1})f(x^{n-1})-\sigma_{n-1}f(x^n)\end{aligned}$$

et comme $\sigma_{n-1}=\sum_{k=1}^{n-1}\lambda_k$,

$$\begin{aligned}&\sum_{k=1}^{n}\sigma_{k-1}f(x^k)-\sum_{k=1}^{n}\sigma_{k-1}f(x^{k-1})\\ &\quad=-\lambda_1f(x^1)-\lambda_2f(x^2)-\lambda_3f(x^3)-\ldots-\lambda_{n-1}f(x^{n-1})+\sum_{k=1}^{n}\lambda_kf(x^n)\\ &\quad=\sum_{k=1}^{n}\lambda_kf(x^n)-\lambda_1f(x^1)-\lambda_2f(x^2)-\ldots-\lambda_{n-1}f(x^{n-1})-\lambda_nf(x^n)\\ &\quad=\sigma_nf(x^n)-\sum_{k=1}^{n}\sigma_{k-1}f(x^{k-1}).\end{aligned}$$

En additionnant (2.11) et (2.9) avec $\lambda_k + \sigma_{k-1} = \sigma_k$, i.e.

$$\sigma_n f(x^n) - \sigma_n f(x) \leq \sum_{k=1}^{n} \sigma_{k-1}\varepsilon_k + H(x, x^0) - H(x, x^n) + \sum_{k=1}^{n} \lambda_k \varepsilon_k \quad \forall x \in C,$$

il suit que

$$f(x^n) - f(x) \leq \frac{1}{\sigma_n}[H(x, x^0) - H(x, x^n)] + \frac{1}{\sigma_n} \sum_{k=1}^{n} \sigma_k \varepsilon_k \quad \forall x \in C, \quad (2.12)$$

ce qui prouve (1) car $H(\cdot, \cdot) \geq 0$.

• Pas 2

Si $\sigma_n \to +\infty$ et $\varepsilon_k \to 0$, alors, en divisant (2.9) par σ_n, i.e.

$$-f(x) + \frac{1}{\sigma_n} \sum_{k=1}^{n} \lambda_k f(x^k) \leq \frac{1}{\sigma_n}[H(x, x^0) - H(x, x^n)] + \frac{1}{\sigma_n} \sum_{k=1}^{n} \lambda_k \varepsilon_k$$

et en invoquant le lemme 2.2.2, nous obtenons de (2.9)

$$\lim \inf_{n \to \infty} f(x^n) \leq \inf\{f(x) \mid x \in C\}.$$

En effet, si nous posons dans le lemme $a_k = f(x^k)$ et $b_n = \frac{1}{\sigma_n} \sum_{k=1}^{n} \lambda_k f(x^k)$, nous obtenons

$$b_n \leq f(x) + \frac{1}{\sigma_n} H(x, x^0) - \frac{1}{\sigma_n} H(x, x^n) + \frac{1}{\sigma_n} \sum_{k=1}^{n} \lambda_k \varepsilon_k.$$

En passant à la limite, nous avons

- $\frac{1}{\sigma_n} H(x, x^0) \to 0$ car $\sigma_n = \sum_{k=1}^{n} \lambda_k \to +\infty$
- $-\frac{1}{\sigma_n} H(x, x^n) \leq 0$ car $H(\cdot, \cdot) \geq 0$
- $\frac{1}{\sigma_n} \sum_{k=1}^{n} \lambda_k \varepsilon_k \to 0$ car, en appliquant le lemme 2.2.2 (2) où nous posons $b_n = \frac{\sum_{k=1}^{n} \lambda_k \varepsilon_k}{\sum_{k=1}^{n} \lambda_k}$ et $a_k = \varepsilon_k$, nous obtenons $\lim a_k = 0$, ce qui implique que $\lim b_n = 0$.

Donc, à la limite, $b_n \leq f(x)$

En appliquant le lemme 2.2.2 (1), nous trouvons le résultat souhaité

$$\liminf_{n\to\infty} f(x^n) \leq \liminf_{n\to\infty} b_n \leq \inf\{f(x) \mid x \in C\}.$$

Comme $f(x^n) \geq \inf\{f(x) \mid x \in \overline{C}\}$, ceci implique que

$$\liminf_{n\to\infty} f(x^n) = \inf\{f(x) \mid x \in \overline{C}\} = f_*.$$

Par (2.10), nous avons

$$0 \leq f(x^k) - f_* \leq f(x^{k-1}) - f_* + \varepsilon_k.$$

Alors, en utilisant le lemme 2.2.1 avec $\beta_k = \varepsilon_k$, $v_k = f(x^k)$ et $\gamma_k = 0$, il suit que la suite $(f(x^k))$ converge vers f_* chaque fois que $\sum_{k=1}^{\infty} \varepsilon_k < \infty$.

- Pas 3

Cas (a)

Par le lemme 2.2.3, nous avons que, si X_* est borné, alors f est coercive sur $\overline{C}$ et comme $(f(x^k)) \to f_*$, il suit que (x^k) est bornée.

De plus, comme f est sci, en passant à la limite et en se rappelant que $(x^k) \subset C$, il suit que chaque valeur d'adhérence est une solution optimale. En effet, par le théorème de Bolzano-Weierstrass, nous savons que, si (x^n) est bornée, alors (x^n) admet une valeur d'adhérence.

Dans notre cas, nous avons $(x^k) \in \{x \mid f(x) \leq v\}$ et $(f(x^k)) \to f_*$. Donc, il existe une sous-suite $(x^{q_k}) \to x^*$ telle que $(f(x^{q_k})) \to f(x^*) = f_*$.

Cas (b)

Ici, nous supposons que $\sum_{k=1}^{n} \lambda_k \varepsilon_k < \infty$ et que $(d, H) \in \mathcal{F}(\overline{C})$.
Alors, (2.8) a lieu pour chaque $x \in \overline{C}$ et en particulier pour $x \in X_*$ de sorte que

$$H(x, x^k) \leq H(x, x^{k-1}) + \lambda_k \varepsilon_k \quad \forall x \in X_*. \tag{2.13}$$

En effet, comme $x \in X_*$, nous avons $f(x) = f_* = \inf\{f(x) \mid x \in \overline{C}\}$.
Donc, $f(x^k) \geq f(x) \quad \forall k$.
Il suit de (2.8) que

$$0 \leq \lambda_k(f(x^k) - f(x)) \leq H(x, x^{k-1}) - H(x, x^k) + \lambda_k \varepsilon_k \quad \forall x \in X_*$$

et nous obtenons (2.13) car $H(\cdot, \cdot) \geq 0$.

En sommant ensuite (2.13) sur $k = 1, \ldots, n$, nous obtenons

$$H(x, x^n) \leq H(x, x^0) + \sum_{k=1}^{\infty} \lambda_k \varepsilon_k.$$

Mais, comme dans ce cas $H(x, \cdot)$ est coercive, la dernière inégalité implique que (x^k) est bornée et donc, comme dans le cas (a), il suit que toutes ses valeurs d'adhérence sont dans X_*. ∎

Comme conséquence de cette analyse, nous obtenons le taux de convergence global estimé pour la version exacte de l'algorithme IPA, i.e. avec ε_k égal à zéro pour tout k.

Corollaire 2.2.1 *Soient* $(d, H) \in \mathcal{F}(\overline{C}), X_* \neq \emptyset$ *et* (x^k) *la suite générée par l'algorithme IPA avec* $\varepsilon_k = 0 \quad \forall k$.

Alors, $f(x^n) - f_* = \mathcal{O}(\sigma_n^{-1}) \quad \forall x \in \overline{C}$.

Preuve

Sous les hypothèses données, le théorème 2.2.1 (1) a lieu pour tout $x \in \overline{C}$ et il suit que $f(x^n) - f_* \leq \dfrac{1}{\sigma_n} H(x^*, x^0)$. ∎

Pour établir la convergence globale de la suite (x^k) vers la solution optimale du problème (P), nous avons besoin de faire plus d'hypothèses sur la distance proximale induite H en imitant le comportement des normes.

Soit $(d, H) \in \mathcal{F}_+(\overline{C}) \subset \mathcal{F}(\overline{C})$ telle que la fonction H satisfait les deux propriétés supplémentaires suivantes :

(1) Pour tout $y \in \overline{C}$ et pour toute suite $(y^k) \subset C$ bornée avec $\lim_{k\to+\infty} H(y, y^k) = 0$, nous avons $\lim_{k\to+\infty} y^k = y$.

(2) Pour tout $y \in \overline{C}$ et pour toute suite $(y^k) \subset C$ convergeant vers y, nous avons $\lim_{k\to+\infty} H(y, y^k) = 0$.

Avec ces hypothèses additionnelles sur la distance proximale induite H, nous obtenons immédiatement que l'algorithme IPA converge globalement vers une solution optimale de (P).

Théorème 2.2.2 *Soit* $(d, H) \in \mathcal{F}_+(\overline{C})$ *et soit* (x^k) *la suite générée par l'algorithme IPA.*

Supposons que l'ensemble optimal X_* *de (P) soit non vide, que* $\sigma_n = \sum_{k=1}^{n} \lambda_k \to \infty$, *que* $\sum_{k=1}^{\infty} \lambda_k \varepsilon_k < \infty$ *et que* $\sum_{k=1}^{\infty} \varepsilon_k < \infty$.

Alors, la suite (x^k) *converge vers une solution optimale de (P).*

Preuve

Soit $x \in X_*$.
Alors, comme $(d, H) \in \mathcal{F}_+(\overline{C})$, par (2.13) avec $\sum_{k=1}^{n} \lambda_k \varepsilon_k < +\infty$ et par le lemme 2.2.1 avec $\beta_k = \lambda_k \varepsilon_k$, $v_k = H(x, x^{k-1})$ et $\gamma_k = 0$, nous obtenons que la suite $(H(x, x^k))$ converge vers un certain $a(x) \in \mathbb{R}$, $\forall x \in X_*$.

Soit x_∞, la limite de la sous-suite (x^{k_l}).
Par le théorème 2.2.1 (3), $x_\infty \in X_*$.

Alors, par l'hypothèse (2), $\lim_{l\to\infty} H(x_\infty, x^{k_l}) = 0$ de telle sorte que $\lim_{k\to\infty} H(x_\infty, x^k) = 0$.

Donc, par l'hypothèse (1), il suit que la suite (x^k) converge vers x_∞. ∎

Notons que nous avons séparé les deux types de résultats de convergence pour mettre en évidence :

- les différences et les rôles joués dans chacune des trois classes $\mathcal{F}_+(\overline{C}) \subset \mathcal{F}(\overline{C}) \subset \mathcal{F}(C)$,
- le fait que la plus grande classe (et la moins demandée) $\mathcal{F}(C)$ fournit déjà des propriétés de convergence raisonnables pour l'algorithme IPA avec des hypothèses minimales sur les données du problème.

Les relations (2.3) et (2.4) définissant l'algorithme IPA peuvent parfois être difficiles à implémenter. Effectivement, à chaque pas, nous devons trouver par un algorithme, en un nombre fini de pas, une solution à ε_k près pour la minimisation de la fonction $\lambda_k f(\cdot) + d(\cdot, x^{k-1})$. Pour venir à bout de cette difficulté, nous considérons ici une variante de la règle d'approximation proposée dans [13] pour les méthodes self-proximales de Bregman.

2.3 Algorithme proximal intérieur avec règle d'approximation

Algorithme Proximal Intérieur avec Règle d'approximation (IPA1)

Soient $(d, H) \in \mathcal{F}(C)$ et $\lambda_* > 0$. Pour chaque $k = 1, 2, \ldots$, soient $\lambda_k \geq \lambda_*, \eta_k > 0$ et ε_k avec $\sum_{k=1}^{\infty} \varepsilon_k < \infty$ et $\sum_{k=1}^{\infty} \eta_k < \infty$.

En partant avec un point $x^0 \in C$, pour tout $k \geq 1$, nous générons les suites $(x^k)_{k=1}^{\infty} \subset C$ et $(e^k)_{k=1}^{\infty} \subset \mathbb{R}^n$ via

$$e^k = \lambda_k g^k + \nabla_1 d(x^k, x^{k-1}) \text{ avec } g^k \in \partial f(x^k), \tag{2.14}$$

où la suite des erreurs (e^k) satisfait les conditions suivantes

$$\|e^k\| \leq \varepsilon_k, \quad \|e^k\| \sup(\|x^k\|, \|x^{k-1}\|) \leq \eta_k. \tag{2.15}$$

Remarque :

Par la proposition 2.1.2, une suite (x^k) donnée par les relations (2.14) et (2.15) existe toujours.

De plus, si $f \in C^1$ sur C, alors, toute méthode de convergence de type gradient fournira une telle suite (x^k) en un nombre fini de pas. Il en est de même avec $f \in C^2$ sur C pour les méthodes de convergence de type Newton.

Théorème 2.3.1 *Soit* $(d, H) \in \mathcal{F}(C)$ *et soit une suite* (x^k) *générée par l'algorithme IPA1.*

Alors, les propriétés suivantes sont valables :

1. *La suite* $(f(x^k))$ *converge vers* f_* .

2. *De plus, supposons que l'ensemble optimal* X_* *soit non vide et considérons les cas suivants :*
   - *(a)* X_* *est borné ;*
   - *(b)* $(d, H) \in \mathcal{F}(\overline{C})$ *;*
   - *(c)* $(d, H) \in \mathcal{F}_+(\overline{C})$.

   Alors, sous (a) *ou* (b) *, la suite* (x^k) *est bornée avec ses valeurs d'adhérence dans* X_* *, tandis que sous* (c) *, la suite* (x^k) *converge vers une solution optimale.*

Preuve

Comme $g^k \in \partial f(x^k)$, en utilisant (2.6) et l'inégalité de Cauchy-Schwarz, nous obtenons, pour tout $x \in C$,

$$\begin{aligned} \lambda_k(f(x^k) - f(x)) &\leq \langle x - x^k, \nabla_1 d(x^k, x^{k-1})\rangle + \langle e^k, x^k - x\rangle \\ &\leq H(x, x^{k-1}) - H(x, x^k) + \tilde{\varepsilon_k}(x), \end{aligned} \tag{2.16}$$

où $\tilde{\varepsilon_k}(x) := \|e^k\|\|x\| + \langle x^k, e^k\rangle$.

En effet, la première inégalité est obtenue comme suit :

$$\begin{aligned} g^k \in \partial f(x^k) \quad \Leftrightarrow \quad & \forall x \in \mathbb{R}^n : f(x) \geq f(x^k) + \langle g^k, x - x^k\rangle \\ & -\langle g^k, x - x^k\rangle \geq f(x^k) - f(x) \\ & -\langle x - x^k, g^k\rangle \geq f(x^k) - f(x) \\ & -\lambda_k\langle x - x^k, g^k\rangle \geq \lambda_k(f(x^k) - f(x)) \\ & \qquad\qquad \text{car } \lambda_k \geq \lambda_* > 0 \end{aligned}$$

$$\begin{aligned}
\langle x - x^k, -\lambda_k g^k \rangle &\geq \lambda_k (f(x^k) - f(x)) \\
\langle x - x^k, \nabla_1 d(x^k, x^{k-1}) - e^k \rangle &\geq \lambda_k (f(x^k) - f(x)) \\
&\text{car } (x^k) \text{ est générée par l'algorithme IPA1} \\
\langle x - x^k, \nabla_1 d(x^k, x^{k-1}) \rangle - \langle x - x^k, e^k \rangle &\geq \lambda_k (f(x^k) - f(x)) \\
\langle x - x^k, \nabla_1 d(x^k, x^{k-1}) \rangle + \langle e^k, x^k - x \rangle &\geq \lambda_k (f(x^k) - f(x)).
\end{aligned}$$

Pour obtenir la seconde inégalité, nous devons montrer que

$$\langle x - x^k, \nabla_1 d(x^k, x^{k-1} \rangle + \langle e^k, x^k - x \rangle \leq H(x, x^{k-1}) - H(x, x^k) + \tilde{\varepsilon_k}.$$

Par (2.6), nous avons $\langle x - x^k, \nabla_1 d(x^k, x^{k-1}) \rangle \leq H(x, x^{k-1}) - H(x, x^k)$.

De plus, nous savons que $\langle e^k, x^k - x \rangle = \langle x^k - x, e^k \rangle = \langle x^k, e^k \rangle - \langle x, e^k \rangle$ et $|\langle x, e^k \rangle| \leq \|x\| \|e^k\|$ par l'inégalité de Cauchy-Schwarz.

Nous avons donc $-\langle x, e^k \rangle < | - \langle x, e^k \rangle| = |\langle x, e^k \rangle| \leq \|x\| \|e^k\|$.

En sommant (2.16) sur $k = 1, \ldots, n$, i.e.

$$\sum_{k=1}^{n} \lambda_k (f(x^k) - f(x)) \leq \sum_{k=1}^{n} [H(x, x^{k-1}) - H(x, x^k)] + \sum_{k=1}^{n} \tilde{\varepsilon_k}(x)$$

où $\sum_{k=1}^{n} [H(x, x^{k-1}) - H(x, x^k)] = H(x, x^0) - H(x, x^n)$,

et en divisant par $\sigma_n = \sum_{k=1}^{n} \lambda_k$, nous obtenons

$$- f(x) + \frac{\sum_{k=1}^{n} \lambda_k f(x^k)}{\sigma_n} \leq \frac{1}{\sigma_n} \left[H(x, x^0) - H(x, x^n) + \sum_{k=1}^{n} \tilde{\varepsilon_k}(x) \right]. \quad (2.17)$$

En prenant alors $x = x^{k-1}$ dans (2.16), i.e.

$$\lambda_k (f(x^k) - f(x^{k-1})) \leq H(x^{k-1}, x^{k-1}) - H(x^{k-1}, x^k) + \tilde{\varepsilon_k}(x^{k-1})$$

et $\alpha_k := |\tilde{\varepsilon_k}(x^{k-1})| \lambda_*^{-1}$, nous obtenons

$$(f(x^k) - f(x^{k-1})) \leq \alpha_k.$$

En effet, nous avons

$$\begin{aligned} f(x^k) - f(x^{k-1}) &\leq \tilde{\varepsilon_k}(x^{k-1})\frac{1}{\lambda_k} && \text{car } H(x,x) = 0 \ \forall x \text{ et } H(\cdot,\cdot) \geq 0 \\ &\leq \tilde{\varepsilon_k}(x^{k-1})\frac{1}{\lambda_*} && \text{car } \lambda_k \geq \lambda_* > 0 \\ &\leq |\tilde{\varepsilon_k}(x^{k-1})|\frac{1}{\lambda_*} =: \alpha_k. \end{aligned}$$

En utilisant (2.15), nous avons $\sum_{k=1}^{\infty} \tilde{\varepsilon_k} < \infty$ et $\sum_{k=1}^{\infty} \alpha_k < \infty$.

En effet, pour obtenir la première inégalité, nous devons repartir de la définition de $\tilde{\varepsilon_k}(x)$:

$$\sum_{k=1}^{\infty} \tilde{\varepsilon_k}(x) = \sum_{k=1}^{\infty} \|e^k\|\|x\| + \sum_{k=1}^{\infty} \langle x^k, e^k \rangle.$$

Or, $\|e^k\| \leq \varepsilon_k$, donc

$$\sum_{k=1}^{\infty} \|e^k\|\|x\| \leq \|x\| \sum_{k=1}^{\infty} \varepsilon_k < \infty \qquad \text{car} \sum_{k=1}^{\infty} \varepsilon_k < \infty.$$

Par l'inégalité de Cauchy-Schwarz et par (2.15), nous avons

$$|\langle x^k, e^k \rangle| \leq \|x^k\|\|e^k\| \leq \eta_k.$$

Donc, $\sum_{k=1}^{\infty} |\langle x^k, e^k \rangle| \leq \sum_{k=1}^{\infty} \eta_k < \infty$.

Or, comme la convergence absolue implique la convergence simple, nous avons $\sum_{k=1}^{\infty} \langle x^k, e^k \rangle < \infty$.

Pour obtenir la seconde inégalité, il faut simplement utiliser la définition de α_k :

$$\sum_{k=1}^{\infty} \alpha_k = \left(\sum_{k=1}^{\infty} |\tilde{\varepsilon_k}(x^{k-1})| \right) \frac{1}{\lambda_*} < \infty$$

car nous venons de montrer que $\sum_{k=1}^{\infty} \tilde{\varepsilon_k} < \infty$.

Donc, en passant à la limite dans (2.17) et en invoquant le lemme 2.2.2 (1), il suit que
$\liminf_{n\to\infty} f(x^n) - f(x) \leq 0 \quad \forall x \in C$ tel que

$$\liminf_{n\to\infty} f(x^n) \leq \inf\{f(x) \mid x \in C\}.$$

En effet, en prenant $b_n = \frac{1}{\sigma_n}\sum_{k=1}^{n} \lambda_k f(x^k)$ et $a_k = f(x^k)$ dans le lemme, nous obtenons par (2.17) et en passant à la limite,

$$\lim_{n\to\infty} b_n \leq f(x) + \lim_{n\to\infty} \frac{1}{\sigma_n} H(x, x^0) - \lim_{n\to\infty} \frac{1}{\sigma_n} H(x, x^n) + \frac{1}{\sigma_n}\sum_{k=1}^{n} \tilde{\varepsilon_k}(x).$$

Or, comme $\sigma_n \to +\infty$, $H(\cdot,\cdot) \geq 0$ et $\sum_{k=1}^{n} \tilde{\varepsilon_k}(x) < \infty$, nous avons $\lim_{n\to\infty} b_n \leq f(x)$.

Enfin, en appliquant le lemme, nous obtenons

$$\liminf_{n\to\infty} a_n = \liminf_{n\to\infty} f(x^n) \leq \liminf_{n\to\infty} b_n \leq f(x)$$

ce qui nous donne le résultat recherché.

A partir d'ici, la preuve peut être terminée en utilisant les mêmes arguments que dans les preuves des théorèmes 2.2.1 et 2.2.2.
Il reste à montrer le point (2) où (a) et (b) suivent les mêmes arguments que le théorème 2.2.1 et (c), ceux du théorème 2.2.2. ∎

2.4 Exemples de distances proximales (d, H)

2.4.1 Distances proximales de Bregman

Dans la plupart des situations, quand nous construisons l'algorithme IPA pour résoudre le problème convexe (P), la distance proximale H induite par d est une distance proximale de Bregman, D_h, générée par un certain noyau convexe h.

Soit $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe, propre et sci avec dom $h \subset \overline{C}$ et dom $\nabla h = C$. Notons que la fonction est strictement convexe sur dom h et C^1 sur int dom $h = C$.

Définissons

$$\begin{aligned} H(x,y) &:= D_h(x,y) \\ &:= \begin{cases} h(x) - [h(y) + \langle \nabla h(y), x - y \rangle] & \forall x \in \mathbb{R}^n, \forall y \in \text{ dom } \nabla h\,; \\ +\infty & \text{autrement.} \end{cases} \end{aligned} \tag{2.18}$$

Notons que, par convexité de h, la distance de Bregman est toujours non négative.

Définition 2.4.1 *Etant donné C un sous-ensemble convexe ouvert de $\mathbb{R}^n$, $h : \overline{C} \to \mathbb{R}$ est appelée fonction de Bregman de zone C si*

1. *h est strictement convexe et continue sur $\overline{C}$;*
2. *h est continûment différentiable sur C ;*
3. *pour tout $x \in C$ et $\alpha \in \mathbb{R}$, l'ensemble*

$$L(x,\alpha) = \{y \in C \mid D_h(x,y) \leq \alpha\}$$

   est borné ;

4. *si (y^k) est une suite de C convergeant vers y, alors*

$$\lim_{k \to \infty} D_h(y, y^k) = 0.$$

Lemme 2.4.1 ***Identité remarquable des trois points***

Soit C un sous-ensemble convexe non vide de $\mathbb{R}^n$.
Soit g une fonction de Bregman sur C.
Pour tout $x, z \in C$ et $y \in \overline{C}$, nous avons

$$D_g(y,x) = D_g(z,x) + D_g(y,z) + \langle \nabla g(x) - \nabla g(z), z - y \rangle.$$

La fonction H admet l'identité des trois points

$$H(c,a) = H(c,b) + H(b,a) + \langle c - b, \nabla_1 H(b,a) \rangle \tag{2.19}$$

$\forall a, b \in C, \forall c \in \text{dom } h$. Cette identité joue un rôle central dans l'analyse de la convergence.

Nous devons considérer deux types de noyaux h différents pour traiter les cas sur C et ceux sur $\overline{C}$.

Le premier type de fonctions, appelées fonctions de Bregman de zone C (voir ci-dessus et [9]), satisfait les conditions suivantes :

(B_1) dom $h = \overline{C}$;

(B_2) (i) $\forall x \in \overline{C}, D_h(x,\cdot)$ est coercive sur int (dom h) ;

(ii) $\forall y \in C, D_h(\cdot, y)$ est coercive ;

(B_3) $\forall y \in$ dom $h, \forall (y^k) \subset$ int (dom h) avec $\lim_{k\to\infty} y^k = y$, nous avons $\lim_{k\to\infty} D_h(y, y^k) = 0$;

(B_4) si (y^k) est une suite bornée sur int (dom h) et $y \in$ dom h telle que $\lim_{k\to\infty} D_h(y, y^k) = 0$, alors $y = \lim_{k\to\infty} y^k$.

Notons que (B_4) est une conséquence directe des trois premières propriétés, ce qui a été prouvé par Kiwiel [16].

Soit $\mathcal{B}$ la classe des noyaux h satisfaisant les propriétés (B_1) à (B_4). Par simplicité, nous considérons ici uniquement le cas où $h \in \mathcal{B}$.

Pour le second type de noyaux, il faut que le noyau convexe h satisfasse aux deux conditions « plus faibles » :

(WB_1) dom $h = C$;

(WB_2) (i) $\forall x \in C, D_h(x, \cdot)$ est coercive sur C ;

(ii) $\forall y \in C, D_h(\cdot, y)$ est coercive.

Nous notons $\mathcal{WB}$ l'ensemble de tels noyaux convexes.

Nous allons à présent donner quelques exemples précis de fonctions de Bregman.

Exemple 3 *Soit* $C = \mathbb{R}^n$ *et considérons la fonction* $h(x) = \|x\|^2$.
Dans ce cas, $D_h(x, y) = \|x - y\|^2$.

En effet, si nous développons à partir de la définition, nous obtenons

$$\begin{aligned} D_h(x,y) &= h(x) - [h(y) + \langle \nabla h(y), x - y\rangle] \\ &= \|x\|^2 - \|y\|^2 - \langle 2y, x - y\rangle \\ &= \|x\|^2 - \|y\|^2 - 2\langle y, x\rangle + 2\langle y, y\rangle \\ &= \|x\|^2 - \|y\|^2 - 2\langle y, x\rangle + 2\|y\|^2 \\ &= \|x\|^2 + \|y\|^2 - 2\langle y, x\rangle \\ &= \|x - y\|^2. \end{aligned}$$

Exemple 4 *Posons* $C = \mathbb{R}^n$ *et considérons la fonction*

$$h(x) = \sum_{j=1}^{n} x^j \log x^j,$$

avec la convention que $0 \log 0 = 0$.
Alors, nous obtenons

$$D_h(x, y) = \sum_{j=1}^{n} \left(x^j \log \frac{x^j}{y^j} + y^j - x^j \right).$$

Notons que cette fonction $D_h(x, y)$ *est la divergence de Kullback-Leibler qui est largement utilisée en statistiques.*

Si nous calculons, par exemple, $D_h(x, y)$ dans le cas où $n = 1$, nous obtenons bien le résultat voulu.

En effet,

$$\begin{aligned}
D_h(x,y) &= x^1 \log x^1 - \left[y^1 \log y^1 + \left\langle \log y^1 + \frac{y^1}{y^1}, x^1 - y^1 \right\rangle \right] \\
&= x^1 \log x^1 - y^1 \log y^1 - \langle \log y^1 + 1, x^1 - y^1 \rangle \\
&= x^1 \log x^1 - y^1 \log y^1 - \langle \log y^1, x^1 - y^1 \rangle - \langle 1, x^1 - y^1 \rangle \\
&= x^1 \log x^1 - y^1 \log y^1 - \langle \log y^1, x^1 \rangle + \langle \log y^1, y^1 \rangle - \langle 1, x^1 - y^1 \rangle \\
&= x^1 \log x^1 - y^1 \log y^1 - x^1 \log y^1 + y^1 \log y^1 - x^1 + y^1 \\
&= x^1 \log x^1 - x^1 \log y^1 - x^1 + y^1 \\
&= x^1 (\log x^1 - \log y^1) - x^1 + y^1 \\
&= x^1 \log \left(\frac{x^1}{y^1} \right) + y^1 - x^1
\end{aligned}$$

Les exemples suivants ne seront pas détaillés, mais nous pouvons faire le même genre de manipulations.

Exemple 5 *Posons* $C = \mathbb{R}^n$ *et considérons la fonction*

$$h(x) = \sum_{j=1}^{n} ((x^j)^\alpha - (x^j)^\beta) \quad \textit{avec } \alpha \geq 1, 0 < \beta < 1.$$

Pour $\alpha = 2, \beta = \frac{1}{2}$, *nous obtenons*

$$D_h(x,y) = \|x - y\|^2 + \sum_{j=1}^{n} \frac{1}{2\sqrt{y^j}} (\sqrt{x^j} - \sqrt{y^j})^2$$

et pour $\alpha = 1, \beta = \frac{1}{2}$, *nous obtenons*

$$D_h(x,y) = \sum_{j=1}^{n} \frac{1}{2\sqrt{y^j}} (\sqrt{x^j} - \sqrt{y^j})^2.$$

Exemple 6 *Supposons que*

$$C = \{x \in \mathbb{R}^n \mid a_j < x^j < b_j, j = 1, \ldots, n\},$$

où, pour chaque j, $-\infty < a_j < b_j < +\infty$.
Considérons la fonction

$$h(x) = \sum_{j=1}^{n} [(x^j - a_j) \log(x^j - a_j) + (b_j - x^j) \log(b_j - x^j)]$$

avec la convention que $0 \log 0 = 0$.
Dans ce cas,

$$D_h(x,y) = \sum_{j=1}^{n} \left[(x^j - a_j) \log \left(\frac{x^j - a_j}{y^j - a_j} \right) + (b_j - x^j) \log \left(\frac{b_j - x^j}{b_j - y^j} \right) \right].$$

Exemple 7 *Soit* $C = \mathbb{R}^n_{++}$. *Les distances proximales séparables de Bregman sont celles qui sont le plus couramment utilisées dans la littérature.*

Soit $\theta : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ *une fonction propre, convexe et sci avec* $(0, +\infty) \subset dom\ \theta \subset [0, +\infty)$ *et telle que* $\theta \in C^2(0, +\infty), \theta''(t) > 0\ \forall t > 0$ *et* $\lim_{t \to 0^+} \theta'(t) = -\infty$.

Nous notons cette classe Θ_0 *si* $\theta(0) < +\infty$ *et* Θ_+ *quand* $\theta(0) = +\infty$.
Notons que θ *est aussi supposée décroissante.*

Etant donné θ *dans chacune des classes, définissons* $h(x) = \sum_{j=1}^{n} \theta(x^j)$ *de telle sorte que* D_h *est séparable.*

Les deux premiers exemples ci-dessous sont des fonctions $\theta \in \Theta_0$, *c'est-à-dire avec dom* $\theta = [0, +\infty)$ *et les deux derniers dans* Θ_+ , *c'est-à-dire avec dom* $\theta = (0, +\infty)$:

- $\theta_1(t) = t\ log\ t$ *(entropie de Shannon);*
- $\theta_2(t) = \dfrac{pt - t^p}{1 - p}$ *avec* $p \in (0, 1)$;
- $\theta_3(t) = -log\ t$ *(entropie de Burg);*
- $\theta_4(t) = t^{-1}$.

Alors, les distances proximales correspondantes D_{h_1} *et* D_{h_2} *appartiennent à* $\mathcal{B}$ *tandis que* D_{h_3} *et* D_{h_4} *appartiennent à* $\mathcal{WB}$.

Des exemples supplémentaires peuvent être trouvés dans [16].

2.4.2 Méthodes self-proximales

L'identité des trois points joue un rôle fondamental dans la convergence des méthodes self-proximales basées sur Bregman, à savoir celles pour lesquelles nous prenons la distance d elle-même comme distance proximale de Bregman. Nous avons donc $d(x,y) = H(x,y) = D_h(x,y)$ avec D_h définie par (2.18).
De plus, quand $h \in \mathcal{B}$ ou $\mathcal{WB}$, les propriétés (P_1), (P_2) et (P_3) ont lieu pour $d = D_h$.

Clairement, $D_h(a,a) = 0 \ \forall a \in C$ de telle sorte que (P_4) a lieu et, comme H est toujours positive, il suit de (2.19) que (2.6) a lieu aussi.
Donc, pour $h \in \mathcal{WB}$, nous avons $(d,H) = (D_h, D_h) \in \mathcal{F}(C)$ et pour $h \in \mathcal{B}$, nous avons $(d,H) = (D_h, D_h) \in \mathcal{F}_+(\overline{C})$.

Quand $C = \mathbb{R}^n$, avec $h(\cdot) = \dfrac{\|\cdot\|^2}{2} \in \mathcal{B}$, alors $D_h(x,y) = \dfrac{\|x-y\|^2}{2}$ et, avec $(d,H) = (D_h, D_h) \in \mathcal{F}_+(\mathbb{R}^n)$, l'algorithme IPA est exactement la méthode proximale classique et les théorèmes (2.2.1) et (2.2.2) couvrent les résultats de convergence habituels (voir, par exemple, [14], [17] et [18]).

Nous allons à présent citer une liste de cas spéciaux intéressants pour la paire (d,H) qui donnent des schémas self-proximaux pour différents types de contraintes.

- **Contraintes non négatives**

Soient $C = \mathbb{R}^n_{++}$ et $\overline{C} = \mathbb{R}^n_+$.
Pour les cas donnés dans l'exemple 7, les algorithmes self-proximaux résultants, à savoir avec $d = H = D_{h_i}$ donnent $(d, D_{h_i} \in \mathcal{F}_+(\overline{C}))$ pour $i = 1,2$ et $(d, D_{h_i} \in \mathcal{F}(C))$ pour $i = 3,4$.

- **Contraintes semi-définies**

Nous notons S^n l'espace linéaire des matrices réelles symétriques pour lesquelles $\langle x,y\rangle := \mathrm{tr}(xy)$ et $\|x\| = \sqrt{\mathrm{tr}(x^2)} \ \forall x,y \in S^n$ où $\mathrm{tr}(x)$ est la trace de la matrice x et $\det(x)$ son déterminant.
Rappelons également quelques autres propriétés de la trace que nous utilise-

rons plus tard :

$$\begin{aligned}
\mathrm{tr}(a+b) &= \mathrm{tr}(a) + \mathrm{tr}(b) && \forall a, b \in S^n, \\
\mathrm{tr}(ka) &= k\mathrm{tr}(a) && \forall a \in S^n, \forall k \text{ scalaire}, \\
\mathrm{tr}(a) &= \mathrm{tr}(a^T) && \forall a \in S^n, \\
\mathrm{tr}(ab) &= \mathrm{tr}(ba) && \forall a, b \in S^n.
\end{aligned}$$

Le cône formé de $n \times n$ matrices semi-définies positives (resp. définies positives) est noté S^n_+ (resp. S^n_{++}).

Posons $C = S^n_{++}$ et $\overline{C} = S^n_+$. Soient

- $h_1 : S^n_+ \to \mathbb{R}, \quad h_1(x) = \mathrm{tr}(x \log x)$;
- $h_3 : S^n_{++} \to \mathbb{R}, \quad h_3(x) = -\mathrm{tr}(\log x) = -\log \det(x)$.

Pour tout $y \in S^n_{++}$, posons

$$\begin{aligned}
d_1(x,y) &= \mathrm{tr}(x\log x - x\log y + y - x) \qquad \text{avec dom } d_1(\cdot, y) = S^n_+, \\
d_3(x,y) &= \mathrm{tr}(-\log x + \log y + xy^{-1}) - n \\
&= -\log\det(xy^{-1}) + \mathrm{tr}(xy^{-1}) - n \text{ avec dom } d_3(\cdot, y) = S^n_{++}.
\end{aligned}$$

En effet, regardons plus en détails le cas de $h_3(x)$:

$$\begin{aligned}
d_3(x,y) &= h_3(x) - h_3(y) - \langle \nabla h_3(y), x - y \rangle \\
&= -\mathrm{tr}(\log x) + \mathrm{tr}(\log y) - \langle -y^{-1}, x - y \rangle \\
&= \mathrm{tr}(-\log x) + \mathrm{tr}(\log y) + \langle y^{-1}, x \rangle - \langle y^{-1}, y \rangle \\
&= \mathrm{tr}(-\log x) + \mathrm{tr}(\log y) + \mathrm{tr}(y^{-1}x) - \mathrm{tr}(y^{-1}y) \\
&= \mathrm{tr}(-\log x) + \mathrm{tr}(\log y) + \mathrm{tr}(xy^{-1}) - tr(I) \\
&= \mathrm{tr}(-\log x + \log y + xy^{-1}) - n
\end{aligned}$$

Nous obtenons le résultat souhaité en appliquant les différentes propriétés de la trace citées ci-dessus.
Nous pouvons évidemment obtenir $d_1(x,y)$ en appliquant le même raisonnement.
Notons aussi que le détail pour obtenir le gradient de $h_3(x)$ peut être trouvé dans [21].

Les distances proximales d_1 et d_3 sont de type Bregman, correspondant respectivement à h_1 et h_3 et ont été proposées par Doljansky et Teboulle dans [12]. A partir de leurs résultats, il est facile de voir que $d_i \in \mathcal{D}(C), i = 1,3$ et avec $H(x,y) = d_i(x,y)$, il suit que $(d_1, H) \in \mathcal{F}(S^n_+)$ et $(d_3, H) \in \mathcal{F}(S^n_{++})$ de telle sorte que nous retrouvons les résultats de [12] à travers le théorème 2.2.1.

Cependant, comme noté dans un contre-exemple dans [12], la propriété (B_3) n'est pas valable pour d_1, ce qui implique qu'elle ne l'est pas pour d_3 non plus. Donc, $(d_i, H) \notin \mathcal{F}(\overline{C})$, $i = 1,3$. En conséquence, le théorème 2.2.2 n'est pas applicable : la convergence globale vers une solution optimale ne peut pas être garantie.

Des résultats similaires peuvent facilement être étendus au cas plus général avec $C = \{x \in \mathbb{R}^m \mid B(x) \in S^n_{++}\}$ supposé non vide, avec $B(x) = \sum_{i=1}^{n} x^i B^i - B^0$, où $B^i \in S^n \ \forall i = 0, \ldots, m$ et avec l'application $x \to \sum_{i=1}^{m} x^i B^i$, en considérant les distances proximales correspondantes

$$D_1(x,y) = d_1(B(x), B(y)) \quad \text{et} \quad D_3(x,y) = d_3(B(x), B(y)).$$

- **Programmation convexe**

Soit $f_i : \mathbb{R}^n \to \mathbb{R}$ des fonctions concaves et C^1 sur $\mathbb{R}^n$ pour tout $i \in [1,m]$.

Nous supposons que la condition de Slater a lieu, c'est-à-dire qu'il existe un certain point $x_0 \in \mathbb{R}^n$ tel que $f_i(x_0) > 0 \ \forall i = 1, \ldots, m$ et que l'ensemble convexe ouvert C est décrit par

$$C = \{x \in \mathbb{R}^n \mid f_i(x) > 0 \ \forall i = 1, \ldots, m\}$$

de telle sorte que, par la condition de Slater, $C \neq \emptyset$ et $\overline{C} = \{x \in \mathbb{R}^n \mid f_i(x) \geq 0 \ \forall i = 1, \ldots, m\}$.

Considérons la classe Θ_+ des fonctions définies dans l'exemple 7 et pour

chaque $\theta \in \Theta_+$, posons

$$h(x) = \begin{cases} \sum_{i=1}^{m} \theta(f_i(x)) & \text{si } x \in C, \\ +\infty & \text{sinon.} \end{cases} \tag{2.20}$$

Evidemment, h est une fonction convexe, propre et sci.

A présent, considérons la distance proximale de Bregman associée à

$$h_\nu(x) := h(x) + \frac{\nu}{2}\|x\|^2 \quad \text{avec } \nu > 0.$$

Alors, nous prenons $d(x,y) = D_{h_\nu}(x,y)$ où D_{h_ν} est la distance de Bregman associée à h_ν.

Grâce à la condition $\nu > 0$, il suit que $h_\nu \in \mathcal{WB}$ et $(d, D_{h_\nu}) \in \mathcal{F}(C)$.

Un cas important et intéressant peut être obtenu en choisissant la fonction de Burg, $\theta_3 = -\log t$. Dans ce cas, nous obtenons

$$d(x,y) = \sum_{i=1}^{m} -\log\frac{f_i(x)}{f_i(y)} + \frac{\langle \nabla f_i(y), x-y\rangle}{f_i(y)} + \frac{\nu}{2}\|x-y\|^2. \tag{2.21}$$

En effet,

$$h_\nu(x) = \begin{cases} \sum_{i=1}^{m} -\log(f_i(x)) + \frac{\nu}{2}\|x\|^2 & \text{si } x \in C \\ +\infty & \text{sinon} \end{cases}$$

Nous pouvons alors calculer la distance de Bregman associée à $h_\nu(x)$.

$$\begin{aligned}
d(x,y) &= h_\nu(x) - h_\nu(y) - \langle \nabla h_\nu(y), x-y\rangle \\
&= \sum_{i=1}^{m} -\log(f_i(x)) + \frac{\nu}{2}\|x\|^2 + \sum_{i=1}^{m}\log(f_i(y)) - \frac{\nu}{2}\|y\|^2 \\
&\qquad - \left\langle -\frac{\nabla f_i(y)}{f_i(y)} + \nu y, x-y \right\rangle \\
&= \sum_{i=1}^{m} -\log\frac{f_i(x)}{f_i(y)} + \frac{\langle \nabla f_i(y), x-y\rangle}{f_i(y)} + \frac{\nu}{2}\|x\|^2 - \frac{\nu}{2}\|y\|^2 \\
&\qquad + \nu\|y\|^2 - \nu\langle y, x\rangle \\
&= \sum_{i=1}^{m} -\log\frac{f_i(x)}{f_i(y)} + \frac{\langle \nabla f_i(y), x-y\rangle}{f_i(y)} + \frac{\nu}{2}\|x\|^2 + \frac{\nu}{2}\|y\|^2 - \nu\langle y, x\rangle \\
&= \sum_{i=1}^{m} -\log\frac{f_i(x)}{f_i(y)} + \frac{\langle \nabla f_i(y), x-y\rangle}{f_i(y)} + \frac{\nu}{2}\|x-y\|^2
\end{aligned}$$

ce qui est bien le résultat voulu.

- **Contraintes du cône du second ordre**

Soit $C = L^n_{++} := \{x \in \mathbb{R}^n \mid x_n > (x_1^2 + \ldots + x_{n-1}^2)\}$ l'intérieur du cône de Lorentz avec sa fermeture notée L^n_+.

Soit J_n une matrice diagonale avec ses $(n-1)$ premiers éléments égaux à -1 et son dernier élément égal à 1.

Nous définissons $h : L^n_{++} \to \mathbb{R}$ par $h(x) = -\log(x^T J_n x)$.

Alors h est une fonction propre, sci et convexe sur dom $h = L^n_{++}$.

Soit $h_\nu(x) = h(x) + \dfrac{\nu\|x\|^2}{2}$.

Alors, comme $\nu > 0$, nous avons $h_\nu \in \mathcal{WB}$ et la distance proximale de Bregman associée à h_ν est donnée par

$$D_{h_\nu}(x,y) = -\log\frac{x^T J_n x}{y^T J_n y} + \frac{2x^T J_n y}{y^T J_n y} - 2 + \frac{\nu}{2}\|x-y\|^2. \tag{2.22}$$

En effet, nous obtenons comme fonction

$$h_\nu(x) = -\log\,(x^T J_n x) + \frac{\nu\|x\|^2}{2}$$

et nous avons la distance proximale de Bregman associée comme suit

$$\begin{aligned}
D_{h_\nu}(x,y) &= h_\nu(x) - h_\nu(y) - \langle \nabla h_\nu(y), x-y\rangle \\
&= -\log(x^T J_n x) + \frac{\nu}{2}\|x\|^2 + \log(y^T J_n y) - \frac{\nu}{2}\|y\|^2 \\
&\quad - \left\langle -\frac{2J_n y}{y^T J_n y} + \nu y, x-y \right\rangle \\
&= -\log\frac{x^T J_n x}{y^T J_n y} + \left\langle \frac{2J_n y}{y^T J_n y}, x-y \right\rangle + \frac{\nu}{2}\|x\|^2 - \frac{\nu}{2}\|y\|^2 \\
&\quad + \nu\|y\|^2 - \nu\langle y, x\rangle \\
&= -\log\frac{x^T J_n x}{y^T J_n y} + 2\left\langle \frac{J_n y}{y^T J_n y}, x \right\rangle - 2\left\langle \frac{J_n y}{y^T J_n y}, y \right\rangle + \frac{\nu}{2}\|x\|^2 \\
&\quad + \frac{\nu}{2}\|y\|^2 - \nu\langle y, x\rangle \\
&= -\log\frac{x^T J_n x}{y^T J_n y} + 2\frac{x^T J_n y}{y^T J_n y} - 2 + \frac{\nu}{2}\|x-y\|^2
\end{aligned}$$

De plus, nous avons $(D_{h_\nu}, D_{h_\nu}) \in \mathcal{F}(L^n_{++})$.

2.4.3 Fonctions proximales basées sur des ϕ-divergences

Ce type de fonction est également un exemple de distances proximales. Dans le cadre de ce travail, nous allons simplement citer les définitions sans les développer.

Ces fonctions montrent que, même dans les cas où l'algorithme IPA n'est pas self-proximal, la distance proximale induite H par le choix de d pour différents types de contraintes sera toujours une distance proximale de Bregman D_h avec un noyau convexe approprié h dans la classe $\mathcal{B}$ ou $\mathcal{WB}$.

Soit $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ une fonction propre convexe et sci telle que dom $\phi \subset \mathbb{R}_+$ et dom $\partial\phi = \mathbb{R}_{++}$. Nous supposons également que ϕ est C^2, strictement convexe et positive sur $\mathbb{R}_{++}$ avec $\phi(1) = \phi'(1) = 0$.

Notons Φ l'ensemble de tels noyaux et Φ_1 la sous-classe de ces noyaux satisfaisant

$$\phi''(1)\left(1 - \frac{1}{t}\right) \leq \phi'(t) \leq \phi''(1)\log t \qquad \forall t > 0. \tag{2.23}$$

L'autre sous-classe de Φ qui est intéressante est notée Φ_2, où (2.23) est remplacée par

$$\phi''(1)\left(1 - \frac{1}{t}\right) \leq \phi'(t) \leq \phi''(1)(t-1) \qquad \forall t > 0. \tag{2.24}$$

Nous définissons ensuite une distance proximale ϕ-divergente correspondant aux classes Φ_r (avec $r = 1, 2$) par

$$d_\phi(x, y) = \sum_{i=1}^{n} y_i^r \phi\left(\frac{x_i}{y_i}\right).$$

Pour tout $\phi \in \Phi$, comme $\arg\min\{\phi(t) \mid t \in \mathbb{R}\} = \{1\}, \phi$ est coercive et donc, il suit que $d_\phi \in \mathcal{D}(C)$ avec $C = \mathbb{R}^n_{++}$.

Chapitre 3

Méthodes du gradient intérieur

3.1 Introduction

Quand $\overline{C} = \mathbb{R}^n$, Correa et Lemaréchal [11] ainsi que Robinson [22] ont remarqué que l'algorithme du point proximal classique peut être vu comme une méthode de descente du sous-gradient approximé.

Cette idée a récemment été étendue par Auslender et Teboulle [1] pour la méthode proximale logarithmique-quadratique, ce qui nous autorise à traiter directement les contraintes d'inégalités linéaires.

Etant donné le cadre développé dans le chapitre 2, nous étendons ces résultats à des contraintes plus générales et avec des classes de distances proximales variables.

Nous allons donner, pour commencer, le principal résultat de convergence. Ensuite, nous présenterons des applications qui nous autorisent à améliorer certaines méthodes connues basées sur le gradient intérieur et à dériver de nouvelles méthodes et des algorithmes de convergence simple pour les problèmes d'optimisation conique.

3.2 Un théorème général de convergence

Pour résoudre le problème

$$\inf\{f(x) \mid x \in \overline{C}\}, \tag{P}$$

considérons l'algorithme suivant :

Algorithme général basé sur le sous-gradient « projeté » (PSA)

Prendre $d \in \mathcal{D}(C)$.

Soient $\lambda_k > 0, \varepsilon_k > 0$ et $m \in (0,1]$.

Soit, pour $k \geq 1$, la suite (x^k, g^k) telle que

$$x^{k-1} \in C, \quad g^{k-1} \in \partial_{\varepsilon_k} f(x^{k-1}), \tag{3.1}$$

$$x^k \in \arg\min\{\lambda_k \langle g^{k-1}, x\rangle + d(x, x^{k-1}) \mid x \in C\}, \tag{3.2}$$

$$f(x^k) \leq f(x^{k-1}) + m(\langle g^{k-1}, x^k - x^{k-1}\rangle - \varepsilon_k). \tag{3.3}$$

Expliquons brièvement pourquoi la suite (x^k) construite par l'algorithme IPA $(\varepsilon_k = 0)$ dans le chapitre 2 convient pour l'algorithme PSA (voir, par exemple, [1], [11] pour plus de détails).

En commençant l'algorithme IPA avec $x^0 \in C$, nous avons $x^k \in C$ et nous pouvons montrer que

$$g^k \in \partial f(x^k) \qquad \Leftrightarrow \qquad g^k \in \partial_{\varepsilon_k^*} f(x^{k-1})$$

avec $\varepsilon_k^* = f(x^{k-1}) - f(x^k) + \langle g^k, x^k - x^{k-1}\rangle \geq 0$.

En effet, $g^k \in \partial_{\varepsilon_k^*} f(x^{k-1}) \quad \Leftrightarrow \quad \forall x \in \text{dom } f$,

$$\begin{aligned} f(x) &\geq f(x^{k-1}) + \langle g^k, x - x^{k-1} \rangle - \varepsilon_k^* \\ &= f(x^{k-1}) + \langle g^k, x - x^{k-1} \rangle - f(x^{k-1}) + f(x^k) - \langle g^k, x^k - x^{k-1} \rangle \\ &= \langle g^k, x - x^{k-1} \rangle + f(x^k) - \langle g^k, x^k - x^{k-1} \rangle \\ &= f(x^k) + \langle g^k, x - x^{k-1} - x^k + x^{k-1} \rangle \\ &= f(x^k) + \langle g^k, x - x^k \rangle \end{aligned}$$

$$\Leftrightarrow \qquad g^k \in \partial f(x^k)\,.$$

Donc, les propriétés (2.3) et (2.4) de l'algorithme IPA sont équivalentes aux propriétés (3.1) et (3.2) de l'algorithme PSA.

En effet, il est évident que (2.3) $\Leftrightarrow$ (3.1).

De plus, par (3.2) et la définition de la distance proximale d, nous savons que le minimum sera toujours à l'intérieur.
Donc, si nous prenons la différentielle de (3.2), nous obtenons

$$0 \in \lambda_k g^{k-1} + \nabla_1 d(x^k, x^{k-1}) + N_C(x^k)$$

ce qui est équivalent à (2.4) car $N_C(x^k)$ est toujours égal à zéro.

Alors, avec $m = 1$ et ε_k^* défini comme ci-dessus, la propriété (3.3) a lieu avec une égalité ce qui nous montre que la suite (x^k) générée par l'algorithme IPA satisfait aux propriétés (3.1), (3.2) et (3.3) de l'algorithme PSA.

En se basant sur ce que nous avons développé dans le chapitre 2, il est maintenant possible d'établir des résultats de convergence de l'algorithme PSA pour différents exemples du triplet $[C, d, H]$ en étendant les résultats de convergence donnés dans [1].

Avant de faire cela, notons d'abord qu'en utilisant les mêmes arguments que dans la preuve de la proposition 2.1.2, nous pouvons voir que l'existence de x^k, appartenant à C, est garantie.

Nous obtenons alors le résultat principal de cette section, à savoir, le théorème général de convergence.

Théorème 3.2.1 *Soit* (x^k) *une suite générée par l'algorithme PSA avec* $(d, H) \in \mathcal{F}(C)$.

Soient $\sigma_n = \sum_{k=1}^{n} \lambda_k$ *et* $\alpha_k = \langle g^{k-1}, x^{k-1} - x^k \rangle$. *Alors,*

1. $\sum_{k=1}^{\infty} \alpha_k < \infty$, $\sum_{k=1}^{\infty} \varepsilon_k < \infty$ *et* $\alpha_k \geq \frac{1}{\lambda_k} H(x^k, x^{k-1}) \geq 0 \quad \forall k \in I\!N$.

2. $\forall z \in C, f(x^n) - f(x) \leq \frac{1}{\sigma_n} \left[H(z, x^0) + \sum_{k=1}^{n} \lambda_k (\alpha_k + \varepsilon_k) \right]$.

3. *La suite* $(f(x^k))$ *est décroissante et converge vers* f_* *quand* $\sigma_n \to \infty$.

4. *Supposons que l'ensemble optimal* X_* *soit non vide et que* $\sigma_n \to \infty$. *Alors, la suite* (x^k) *est bornée avec toutes ses valeurs d'adhérence dans* X_* *sous une des conditions suivantes :*

   (a) X_* *est borné*

   (b) $(d, H) \in \mathcal{F}(\overline{C})$ *et* $\sum_{k=1}^{\infty} \lambda_k \varepsilon_k < +\infty$

   (et c'est particulièrement vrai si (λ_k) *est borné supérieurement).*

   De plus, si $(d, H) \in \mathcal{F}_+(\overline{C})$, *alors* (x^k) *converge vers une solution optimale de (P).*

Preuve

• Pas 1 :

Par les conditions d'optimalité, la propriété (3.2) équivaut à

$$\lambda_k g^{k-1} + \nabla_1 d(x^k, x^{k-1}) = 0.$$

Comme $H(\cdot, \cdot) \geq 0$ et $H(a, a) = 0$, par (2.6) avec $c = a = x^{k-1}$, $b = x^k$, i.e.

$$\langle x^{k-1} - x^k, \nabla_1 d(x^k, x^{k-1}) \rangle \leq H(x^{k-1}, x^{k-1}) - H(x^{k-1}, x^k) \quad \forall x^{k-1} \in C, \quad (*)$$

nous obtenons

$$\lambda_k \alpha_k = \langle \nabla_1 d(x^k, x^{k-1}), x^k - x^{k-1} \rangle \geq H(x^{k-1}, x^k) \geq 0.$$

En effet, par définition de α_k, nous avons

$$\begin{aligned}\lambda_k\alpha_k &= \lambda_k\langle g^{k-1}, x^{k-1}-x^k\rangle \\ &= \langle\lambda_k g^{k-1}, x^{k-1}-x^k\rangle \\ &= \langle-\nabla_1 d(x^k,x^{k-1}), x^{k-1}-x^k\rangle \\ &= \langle\nabla_1 d(x^k,x^{k-1}), x^k-x^{k-1}\rangle\end{aligned}$$

ce qui donne la première égalité.

Il nous reste à montrer que $\lambda_k\alpha_k \geq H(x^{k-1},x^k)$.

Par $(*)$, nous avons

$$\begin{aligned}&\langle x^{k-1}-x^k, \nabla_1 d(x^k,x^{k-1})\rangle \leq -H(x^{k-1},x^k) \\ \Leftrightarrow\quad &\langle x^{k-1}-x^k, \nabla_1 d(x^k,x^{k-1})\rangle \geq H(x^{k-1},x^k) \\ \Leftrightarrow\quad &\langle\nabla_1 d(x^k,x^{k-1}), x^k-x^{k-1}\rangle \geq H(x^{k-1},x^k).\end{aligned}$$

De plus, par (3.3), nous obtenons

$$m(\alpha_k+\varepsilon_k) \leq f(x^{k-1}) - f(x^k), \tag{3.4}$$

ce qui montre que $(f(x^k))$ est décroissante.

Ensuite, en sommant sur $k=1,\ldots,n$ dans (3.4), nous obtenons

$$m\sum_{k=1}^{n}(\alpha_k+\varepsilon_k) \leq f(x^0)-f(x^n) \leq f(x^0)-f_* \tag{3.5}$$

car $x^n \in C$ et $f_* = \inf\{f(x) \mid x\in C\}$,
ce qui prouve (1).

- <u>Pas 2</u> :

Comme $\sigma_n = \sum_{k=1}^{n}\lambda_k$, avec $\sigma_k = \lambda_k + \sigma_{k-1}$ $(\sigma_0=0)$, en multipliant (3.4) par σ_{k-1}, i.e.

$$\sigma_{k-1}m(\alpha_k+\varepsilon_k) \leq \sigma_{k-1}f(x^{k-1}) - \sigma_{k-1}f(x^k)$$

et en sommant sur $k=1,\ldots,n$, nous obtenons

$$\sum_{k=1}^{n}[(\sigma_k-\lambda_k)f(x^k) - \sigma_{k-1}f(x^{k-1})] \leq 0. \tag{3.6}$$

En effet, nous avons

$$\sum_{k=1}^{n} \sigma_{k-1} m(\alpha_k + \varepsilon_k) \leq \sum_{k=1}^{n} [\sigma_{k-1} f(x^{k-1}) - \sigma_{k-1} f(x^k)]$$

or, $\sigma_k = \lambda_k + \sigma_{k-1}$ et $m(\alpha_k + \varepsilon_k) \leq f(x^{k-1}) - f(x^k)$,
donc,

$$\sum_{k=1}^{n} \sigma_{k-1}(f(x^{k-1}) - f(x^k)) \leq \sum_{k=1}^{n} \sigma_{k-1}[f(x^{k-1}) - f(x^k)]$$

et comme $\sigma_k \geq 0$ et $(f(x^k))$ est décroissante, nous avons

$$\sum_{k=1}^{n} \sigma_{k-1}[f(x^{k-1}) - f(x^k)] \geq 0.$$

D'où,

$$\sum_{k=1}^{n} \sigma_{k-1}[f(x^k) - f(x^{k-1})] \leq 0$$

et finalement

$$\sum_{k=1}^{n} [(\sigma_k - \lambda_k) f(x^k) - \sigma_{k-1} f(x^{k-1})] \leq 0.$$

L'inégalité (3.6) se réduit à

$$\sigma_n f(x^n) - \sum_{k=1}^{n} \lambda_k f(x^k) \leq 0. \tag{3.7}$$

Maintenant, comme $g^{k-1} \in \partial_{\varepsilon_k} f(x^{k-1})$, pour tout $z \in C$, nous avons :

$$\begin{aligned} f(z) - f(x^{k-1}) + \varepsilon_k &\geq \langle g^{k-1}, z - x^{k-1} \rangle \\ &= \langle g^{k-1}, z - x^k \rangle + \langle g^{k-1}, x^k - x^{k-1} \rangle \\ &= -\frac{1}{\lambda_k} \langle z - x^k, \nabla_1 d(x^k, x^{k-1}) \rangle - \alpha_k \\ &\geq \frac{1}{\lambda_k} [H(z, x^k) - H(z, x^{k-1})] - \alpha_k \end{aligned}$$

où nous obtenons la deuxième égalité car $\lambda_k g^{k-1} + \nabla_1 d(x^k, x^{k-1}) = 0$ et $-\alpha_k = \langle g^{k-1}, x^k - x^{k-1} \rangle$ ainsi que la dernière inégalité par (2.6) avec $b = x^k$ et $a = x^{k-1}$, i.e.

$$\begin{aligned} -\frac{1}{\lambda_k} \langle z - x^k, \nabla_1 d(x^k, x^{k-1}) \rangle - \alpha_k &\geq -\frac{1}{\lambda_k} [H(z, x^{k-1}) - H(z, x^k)] - \alpha_k \\ &= \frac{1}{\lambda_k} [H(z, x^k) - H(z, x^{k-1})] - \alpha_k. \end{aligned}$$

Comme $f(x^k) \leq f(x^{k-1})$, il suit que

$$\lambda_k(f(x^k) - f(x)) \leq H(z, x^{k-1}) - H(z, x^k) + \lambda_k(\alpha_k + \varepsilon_k). \tag{3.8}$$

En effet, nous avons

$$f(z) - f(x^{k-1}) \geq \frac{1}{\lambda_k}[H(z, x^k) - H(z, x^{k-1})] - \alpha_k - \varepsilon_k$$

$\Leftrightarrow$

$$\lambda_k(f(z) - f(x^{k-1})) \geq H(z, x^k) - H(z, x^{k-1}) - \lambda_k(\alpha_k + \varepsilon_k)$$

ce qui est équivalent au résultat souhaité.

En sommant (3.8) sur $k = 1, \ldots, n$, nous obtenons

$$-\sigma_n f(z) + \sum_{k=1}^{n} \lambda_k f(x^k) \leq H(z, x^0) - H(z, x^n) + \sum_{k=1}^{n} \lambda_k(\alpha_k + \varepsilon_k).$$

En additionnant cette inégalité à (3.7) et en divisant par σ_n, nous obtenons

$$f(x^n) - f(z) \leq \frac{H(z, x^0)}{\sigma_n} + \sum_{k=1}^{n} \frac{\lambda_k(\alpha_k + \varepsilon_k)}{\sigma_n} \quad \forall z \in C. \tag{3.9}$$

Ce qui prouve (2).

- <u>Pas 3</u> :

Supposons que $\sigma_n \to \infty$.
Comme les suites (α_k) et (ε_k) convergent vers zéro, en passant à la limite dans (3.9), i.e.

$$\begin{aligned}
\lim_{n\to\infty} f(x^n) &\leq \lim_{n\to\infty} \left[f(z) + \frac{H(z, x^0)}{\sigma_n} + \sum_{k=1}^{n} \frac{\lambda_k(\alpha_k + \varepsilon_k)}{\sigma_n} \right] \quad \forall x \in C \\
&= f(z) + \lim_{n\to\infty} \frac{1}{\sigma_n} H(z, x^0) + \lim_{n\to\infty} \frac{1}{\sigma_n} \sum_{k=1}^{n} \lambda_k(\alpha_k + \varepsilon_k) \\
&= f(z) + \lim_{n\to\infty} \frac{1}{\sigma_n} \sum_{k=1}^{n} \lambda_k(\alpha_k + \varepsilon_k)
\end{aligned}$$

et par le lemme 2.2.2 (2) avec $b_n = \dfrac{1}{\sigma_n} \displaystyle\sum_{k=1}^{n} \lambda_k(\alpha_k + \varepsilon_k)$ et $a_n = \alpha_k + \varepsilon_k$, comme $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} (\alpha_n + \varepsilon_n) = 0$, nous avons

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} \frac{1}{\sigma_n} \sum_{k=1}^{n} \lambda_k(\alpha_k + \varepsilon_k) = 0.$$

Donc, $\forall z \in C, \lim_{n\to\infty} f(x^n) \leq f(z)$.

Nous obtenons donc finalement que

$$\lim_{n\to\infty} f(x^n) = \lim_{n\to\infty} \sup f(x^n) \leq \inf\{f(x) \mid x \in C\} = f_*$$

et (3) est prouvé.

- Pas 4 :

Pour prouver la fin du théorème 3.2.1, nous utilisons les mêmes arguments que dans les théorèmes 2.2.1 et 2.2.2. ∎

Nous obtenons le corollaire suivant en utilisant le point (2) du théorème 3.2.1 ainsi que la propriété (3.5).

Corollaire 3.2.1 *Soit* $(d, H) \in \mathcal{F}(\overline{C})$ *et soit* (x^k) *une suite produite par l'algorithme PSA.*

Supposons que l'ensemble optimal X_* *soit non vide et que* $0 < \lambda_* \leq \lambda_k \leq \lambda^*$.

Alors, nous avons l'estimation globale suivante $f(x^n) - f_* = \mathcal{O}(n^{-1})$.

Preuve

Par le théorème 3.2.1 (2), nous avons

$$\forall z \in C, \quad f(x^n) - f(z) \leq \frac{1}{\sigma_n}\left[H(z, x^0) + \sum_{k=1}^{n} \lambda_k(\alpha_k + \varepsilon_k)\right].$$

Comme $X_* \neq \emptyset$, supposons qu'il existe $x^* \in X_*$ tel que $f_* = f(x^*) = \inf\{f(x) \mid x \in \overline{C}\}$.

Donc, comme $x^* \in C$,

$$\begin{aligned} f(x^n) - f(x^*) &\leq \frac{1}{\sigma_n}\left[H(x^*, x^0) + \sum_{k=1}^{n} \lambda_k(\alpha_k + \varepsilon_k)\right] \\ &= \underbrace{\frac{1}{\sigma_n} H(x^*, x^0)}_{(I)} + \underbrace{\frac{1}{\sigma_n}\sum_{k=1}^{n} \lambda_k(\alpha_k + \varepsilon_k)}_{(II)}. \end{aligned}$$

Regardons maintenant chacun des termes séparément.

Pour le premier terme,

$$(I) = \frac{1}{\sigma_n} H(x^*, x^0) = \frac{H(x^*, x^0)}{\sum_{k=1}^{n} \lambda_k} \leq \frac{H(x^*, x^0)}{n\lambda_*}$$

car $\sigma_n = \lambda_1 + \lambda_2 + \ldots + \lambda_n \geq n\lambda_*$.

Pour le second terme, nous avons

$$\frac{\lambda_k(\alpha_k + \varepsilon_k)}{\sigma_n} = \frac{\lambda_k(\alpha_k + \varepsilon_k)}{\sum_{k=1}^{n} \lambda_k} \leq \frac{\lambda^*(\alpha_k + \varepsilon_k)}{n\lambda_*}.$$

Donc,

$$(II) = \sum_{k=1}^{n} \frac{\lambda_k(\alpha_k + \varepsilon_k)}{\sigma_n} \leq \sum_{k=1}^{n} \frac{\lambda^*(\alpha_k + \varepsilon_k)}{n\lambda_*} = \frac{\lambda^*}{n\lambda_*} \sum_{k=1}^{n} (\alpha_k + \varepsilon_k).$$

Finalement, en sommant les bornes obtenues pour (I) et (II), nous avons

$$f(x^n) - f(x^*) = f(x^n) - f_* \leq \frac{H(x^*, x^0)}{n\lambda_*} + \frac{\lambda^*}{n\lambda_*} \sum_{k=1}^{n} (\alpha_k + \varepsilon_k)$$

$$\Leftrightarrow \qquad f(x^n) - f_* \leq \frac{1}{n} \frac{H(x^*, x^0)}{\lambda_*}$$

et nous obtenons le résultat souhaité

$$f(x^n) - f_* = \mathcal{O}(n^{-1}).$$

■

3.3 Optimisation conique : Méthodes du gradient intérieur avec une distance proximale fortement convexe

3.3.1 Préliminaires

A partir de maintenant, nous considérons le problème suivant :

$$\inf\{f(x) \mid x \in \overline{C} \cap \mathcal{V}\}, \tag{M}$$

où $\mathcal{V} = \{x : Ax = b\}$ avec $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$, $n \geq m$ et $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe et sci.

Supposons qu'il existe $x^0 \in \text{dom } f \cap C$ tel que $Ax^0 = b$.

Quand $\overline{C}$ est un cône convexe, le problème (M) est le problème d'optimisation conique standard (voir, par exemple, [20]) et quand $\mathcal{V} = \mathbb{R}^n$, il s'agit simplement d'un problème d'optimisation conique pur.

Dans les sous-sections suivantes, nous supposons aussi que f est continûment différentiable avec ∇f lipschitzien sur $C \cap \mathcal{V}$ et de constante de Lipschitz L, i.e. $\exists\, L > 0$ telle que

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \forall x, y \in C \cap \mathcal{V}. \tag{3.10}$$

Nous considérons maintenant $(d, H) \in \mathcal{F}(C)$ tel que d satisfait aux propriétés suivantes :

(1) $\exists\, \sigma > 0 : \forall y \in C \cap \mathcal{V}, d(\cdot, y)$ est σ-fortement convexe sur $C \cap \mathcal{V}$, i.e.

$$\langle \nabla_1 d(x^1, y) - \nabla_1 d(x^2, y), x^1 - x^2 \rangle \geq \sigma \|x^1 - x^2\|^2 \tag{3.11}$$

$\forall x^1, x^2 \in C \cap \mathcal{V}$ et pour une certaine norme $\|.\|$ dans $\mathbb{R}^n$.

(2) $\forall y \in C \cap \mathcal{V}, d(., y)$ est C^2 sur C avec la fonction hessienne notée $\nabla_1^2 d(., y)$.

Donc, avec les mêmes arguments que ceux donnés dans la proposition 2.1.2, il suit que, pour chaque $x \in C \cap \mathcal{V}$, pour chaque $v \in \mathbb{R}^n$, il existe un point unique (par forte convexité) $u(v,x) \in C \cap \mathcal{V}$ tel que

$$u(v,x) = \arg\min\{\langle v, z\rangle + d(z,x) \mid z \in \mathcal{V}\}. \tag{3.12}$$

En effet, le résultat découle directement du théorème 1.3.1.

Donc, par les conditions d'optimalité du problème convexe (3.12),
$\exists\ \mu := \mu(v,x) \in \mathbb{R}^m$ tel que

$$v + A^T\mu + \nabla_1 d(u(v,x),x) = 0, \qquad Au(v,x) = b. \tag{3.13}$$

En effet, si nous regardons les conditions d'optimalité dans le cas des contraintes d'égalité, le lagrangien s'exprime par

$$L(z,\mu) := \langle v, z\rangle + d(z,x) + Az - b$$

et, en dérivant ce dernier par rapport à z, nous obtenons bien le résultat.

Clairement, le problème (M) peut être formulé de façon équivalente à un problème de la forme de (P) comme suit :

$$f_* = \min\{f_0(x) \mid x \in \overline{C}\} \text{ avec } f_0 = f + \delta_{\mathcal{V}}.$$

Définissons $\mathcal{V}_0 = \{x \mid Ax = 0\}$.

Notons que, pour tout $w \in \mathcal{V}$, nous avons $f(w) = f_0(w)$. En effet

$$f_0(w) = f(w) + \delta_{\mathcal{V}}(w) = f(w) + \begin{cases} 0 & \text{si } w \in \mathcal{V}\ ; \\ +\infty & \text{sinon}\ ; \end{cases}$$

et comme nous savons que $w \in \mathcal{V}$, le résultat suit.

Nous avons également

$$\gamma(\eta,x) := (\nabla f(x) + A^T\eta) \in \partial f_0(x) \qquad \forall x \in C \cap \mathcal{V}, \forall \eta \in \mathbb{R}^m. \tag{3.14}$$

En effet, pour tout $z, x \in \mathcal{V}$, nous avons $z - x \in \mathcal{V}_0$ car $A(z-x) = Az - Ax = b - b = 0$ et donc, pour tout $\eta \in \mathbb{R}^m$,

$$\begin{aligned} f_0(z) &= f(z) \text{ car } z \in \mathcal{V}_0 \\ &\geq f(x) + \langle \nabla f(x), z - x \rangle \\ &= f_0(x) + \langle \nabla f(x) + A^T\eta, z - x \rangle \\ &= f_0(x) + \langle \gamma(\eta, x), z - x \rangle . \end{aligned}$$

où l'inégalité découle de la définition du sous-différentiel et où la dernière égalité vient du fait que $x \in \mathcal{V}$ et parce que nous avons

$$\begin{aligned} \langle \nabla f(x) + A^T\mu, z - x \rangle &= \langle \nabla f(x), z - x \rangle + \langle A^T\mu, z - x \rangle \\ &= \langle \nabla f(x), z - x \rangle + \langle \mu, A(z - x) \rangle \\ &= \langle \nabla f(x), z - x \rangle \end{aligned}$$

car z et $x \in \mathcal{V}$.

Comme pour $z \notin \mathcal{V}$, cette inégalité a lieu car $f_0(z) = +\infty$, la propriété (3.14) est vérifiée.

3.3.2 Algorithmes

Nous pouvons à présent proposer, pour résoudre le problème (M), l'itération de base de l'algorithme.

Etant donné une règle de longueur de pas pour choisir λ_k à chaque pas k, en commençant avec un point $x^0 \in C \cap \mathcal{V}$, nous générons itérativement la suite $x^k \in C \cap \mathcal{V}$ par la relation

$$x^k = u(\lambda_k \nabla f(x^{k-1}), x^{k-1}). \tag{3.15}$$

Comme conséquence de la discussion ci-dessus, les relations (3.1) et (3.2) sont satisfaites avec f remplacé par f_0, $\varepsilon_k = 0$ et

$$g^{k-1} = \gamma\left(\frac{\mu(\lambda_k \nabla f(x^{k-1}), x^{k-1})}{\lambda_k}, x^{k-1}\right) \in \partial f_0(x^{k-1}). \tag{3.16}$$

En effet, par (3.15), nous avons $x^{k-1} = u(\lambda_k \nabla f(x^{k-2}), x^{k-2})$ où $x^{k-1} \in C \cap \mathcal{V}$ et donc $x^{k-1} \in C$.
De plus, comme $\varepsilon_k = 0$ et $f = f_0$, nous avons bien $g^{k-1} \in \partial f_0(x^{k-1})$. Il suit donc que (3.1) est satisfaite.

Il nous reste à montrer que (3.2) est aussi satisfaite.
Par (3.12) et (3.15), nous avons

$$\begin{aligned} x^k &= u(\lambda_k \nabla f(x^{k-1}), x^{k-1}) \\ &= \arg\min\{\langle \lambda_k \nabla f(x^{k-1}), z\rangle + d(z, x^{k-1}) \mid z \in \mathcal{V}\} \\ &= \arg\min\{\lambda_k \langle \nabla f(x^{k-1}), z\rangle + d(z, x^{k-1}) \mid z \in \mathcal{V}\}. \end{aligned}$$

Par (3.14), nous avons

$$g^{k-1} = \nabla f(x^{k-1}) + A^T \frac{\mu(\lambda_k \nabla f(x^{k-1}), x^{k-1})}{\lambda_k} =: \nabla f(x^{k-1}) + \frac{A^T \mu}{\lambda_k}.$$

Il suffit alors de montrer que

$$\lambda_k \langle g^{k-1}, z\rangle = \lambda_k \langle \nabla f(x^{k-1}), z\rangle \quad \forall z \in C \cap \mathcal{V}.$$

En effet,

$$\begin{aligned} \lambda_k \langle g^{k-1}, z\rangle &= \lambda_k \left\langle \nabla f(x^{k-1}) + \frac{A^T \mu}{\lambda_k}, z \right\rangle \\ &= \lambda_k \langle \nabla f(x^{k-1}), z\rangle + \langle A^T \mu, z\rangle \\ &= \lambda_k \langle \nabla f(x^{k-1}), z\rangle \end{aligned}$$

car, en prenant $z = x - y$ (avec $x, y \in C \cap \mathcal{V}$), nous obtenons que $z \in \mathcal{V}_0$ et donc

$$\langle A^T \mu, z\rangle = \langle \mu, Az\rangle = \langle \mu, 0\rangle = 0.$$

Nous proposons maintenant deux règles de longueur de pas et, pour chaque règle, nous allons montrer que l'inégalité (3.3) a lieu et que $\sum_{k=1}^{\infty} \lambda_k = \infty$.

Par conséquent, nous serons capables d'appliquer le théorème 3.2.1 et de combiner deux algorithmes du gradient intérieur convergents, ce qui étend

les résultats de Auslender et Teboulle [1].

Algorithme 1 : Règle de longueur de pas constant

Soit $\varepsilon \in]0,1[$.

Posons $\lambda^* := 2\varepsilon\sigma L^{-1}$ et $\lambda_* \in (0, \lambda^*)$.

Commencer à partir d'un point $x^0 \in C \cap \mathcal{V}$ et générer une suite $(x^k) \in C \cap \mathcal{V}$ comme suit :

- Si $\nabla f(x^{k-1}) \in \mathcal{V}_0^{\perp}$, STOP.
- Sinon, calculer

$$x^k = x^k(\lambda_k) := u(\lambda_k \nabla f(x^{k-1}), x^{k-1}) \tag{3.17}$$

avec $\lambda_k \in (\lambda_*, \lambda^*]$.

Citons à présent le lemme de descente que nous utiliserons par après dans les preuves de plusieurs théorèmes.

Lemme 3.3.1 ***Lemme de descente***

Si $f : \mathbb{R}^n \to \mathbb{R}$ *est continûment différentiable et satisfait à la propriété suivante*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \qquad \forall x, y \in \mathbb{R}^n$$

où L *est un scalaire, alors*

$$f(x+y) \leq f(x) + y^T \nabla f(x) + \frac{L}{2}\|y\|^2.$$

Preuve

Soient t un paramètre scalaire et $g(t) = f(x + ty)$.

Alors,

$$\left(\frac{dg}{dt}\right)(t) = y^T \nabla f(x+ty).$$

Nous obtenons

$$\begin{aligned}
f(x+y) - f(x) &= g(1) - g(0) \\
&= \int_0^1 \frac{dg}{dt}(t)\ dt \\
&= \int_0^1 y^T \nabla f(x+ty)\ dt \\
&= \int_0^1 (y^T \nabla f(x+ty) + y^T \nabla f(x) - y^T \nabla f(x))\ dt \\
&\leq \int_0^1 y^T \nabla f(x)\ dt + \left| \int_0^1 y^T (\nabla f(x+ty) - \nabla f(x))\ dt \right| \\
&\leq \int_0^1 y^T \nabla f(x)\ dt + \int_0^1 \|y\| \|\nabla f(x+ty) - \nabla f(x)\|\ dt \\
&\leq \int_0^1 y^T \nabla f(x)\ dt + \|y\| \int_0^1 L\|x - ty + x\|\ dt \\
&\leq y^T \nabla f(x) + \|y\| \int_0^1 Lt\|y\|\ dt \\
&= y^T \nabla f(x) + L\|y\|^2 \int_0^1 t\ dt \\
&= y^T \nabla f(x) + L\|y\|^2 \left[\frac{t^2}{2}\right]_0^1 \\
&= y^T \nabla f(x) + \frac{L}{2}\|y\|^2.
\end{aligned}$$

■

Théorème 3.3.1 *Soit* (x^k) *une suite produite par l'algorithme 1.*

Si, au pas k, *nous avons* $\nabla f(x^{k-1}) \in \mathcal{V}_0^{\perp}$, *alors* x^{k-1} *est une solution optimale.*

Sinon, la suite $(f(x^k))$ *n'est pas croissante et converge vers* f_*.

De plus, supposons que l'ensemble optimal X_* *soit non vide, alors*

1. *si* X_* *est borné, la suite* (x^k) *est bornée avec toutes ses valeurs d'adhérence dans* X ;
2. *si* $(d, H) \in \mathcal{F}_+(\overline{C})$, *la suite* (x^k) *converge vers une solution optimale de* (P).

Preuve

Si $\nabla f(x^{k-1}) \in \mathcal{V}_0^{\perp}$, comme $x^{k-1} \in C \cap \mathcal{V}$, alors, par les conditions d'optimalité (3.13), x^{k-1} est une solution optimale. Ce résultat est obtenu par l'algorithme 1.

Supposons maintenant que $\nabla f(x^{k-1}) \notin \mathcal{V}_0^{\perp}$.
Comme $\lambda_k \geq \lambda_*$, alors $\sigma_n = \sum_{k=1}^{n} \lambda_k \to \infty$.
En effet, par l'algorithme 1, nous savons que $\lambda_* \in (0, \lambda^*)$ où $\lambda^* := 2\varepsilon\sigma L^{-1}, \varepsilon \in]0,1[$.

Nous obtenons donc $0 < \lambda_* < \lambda^* = 2\varepsilon\sigma L^{-1} \leq \lambda_k$.

En prenant la somme, nous avons

$$0 < \sum_{k=1}^{n} \lambda^* = n\lambda^* \leq \sum_{k=1}^{n} \lambda_k$$

et en passant à la limite, nous obtenons le résultat.

Donc, il reste à prouver (3.3) et le résultat suivra comme une conséquence directe du théorème 3.2.1.

Comme ∇f est Lipschitz, par le lemme de descente, nous avons

$$f(x^k) \leq f(x^{k-1}) + \langle \nabla f(x^{k-1}), x^k - x^{k-1} \rangle + \frac{L}{2}\|x^k - x^{k-1}\|^2. \tag{3.18}$$

Remarquons que

$$(x^k - x^{k-1}) \in \mathcal{V}_0. \tag{3.19}$$

En effet, comme x^k et $x^{k-1} \in \mathcal{V}$ (car $(x^k) \subset C \cap \mathcal{V}$ est générée par l'algorithme 1), nous avons

$$A(x^k - x^{k-1}) = Ax^k - Ax^{k-1} = b - b = 0.$$

Alors, en utilisant (3.11) avec $x_1 = y = x^{k-1} \in C \cap \mathcal{V}$, $x_2 = u(v, x^{k-1}) \in C \cap \mathcal{V}$ et $v = \lambda_k \nabla f(x^{k-1})$, (3.13) et g^{k-1} défini comme dans (3.16) (et en se rappelant que $\nabla_1 d(y, y) = 0$), il suit que

$$\lambda_k \langle g^{k-1}, x^{k-1} - x^k \rangle = \lambda_k \langle \nabla f(x^{k-1}), x^{k-1} - x^k \rangle \geq \sigma \|x^k - x^{k-1}\|^2. \tag{3.20}$$

En effet,

$$g^{k-1} = \gamma\left(\frac{\mu(\lambda_k \nabla f(x^{k-1}), x^{k-1})}{\lambda_k}, x^{k-1}\right).$$

Or, $\gamma(\eta, x) := \nabla f(x) + A^T \eta$, donc

$$g^{k-1} = \nabla f(x^{k-1}) + A^T \frac{\mu(\lambda_k \nabla f(x^{k-1}), x^{k-1})}{\lambda_k} = \nabla f(x^{k-1}) + \frac{A^T \mu}{\lambda_k}$$

$$\Leftrightarrow \qquad A^T \mu = \lambda_k (g^{k-1} - \nabla f(x^{k-1})). \tag{$*$}$$

Par (3.13), nous avons

$$A^T \mu = -\nabla_1 d(u(v, x), x) - v. \tag{$**$}$$

En égalant $(*)$ et $(**)$, nous obtenons

$$\lambda_k (g^{k-1} - \nabla f(x^{k-1})) = -\nabla_1 d(u(v, x), x) - v. \tag{3.21}$$

Par (3.15), nous avons $u(v, x^{k-1}) = x^k$, ce qui donne, en insérant ce résultat dans (3.11),

$$\langle -\nabla_1 d(x^k, x^{k-1}), x^{k-1} - x^k \rangle \geq \sigma \|x^{k-1} - x^k\|^2$$

$$\Leftrightarrow \qquad \langle -\nabla_1 d(x^k, x^{k-1}), x^{k-1} - x^k \rangle \geq \sigma \|x^k - x^{k-1}\|^2.$$

Il reste donc à montrer que

$$\begin{aligned}\lambda_k\langle g^{k-1}, x^{k-1}-x^k\rangle &= \lambda_k\langle \nabla f(x^{k-1}), x^{k-1}-x^k\rangle \\ &= \langle -\nabla_1 d(x^k, x^{k-1}), x^{k-1}-x^k\rangle.\end{aligned}$$

Par (3.21), nous obtenons la deuxième égalité car $v = \lambda_k \nabla f(x^{k-1})$.
Ensuite, comme $g^{k-1} = \nabla f(x^{k-1}) + \dfrac{A^T\mu}{\lambda_k}$, nous avons

$$\lambda_k\langle g^{k-1}, x^{k-1}-x^k\rangle = \lambda_k\langle \nabla f(x^{k-1}), x^{k-1}-x^k\rangle + \langle A^T\mu, x^{k-1}-x^k\rangle,$$

où

$$\langle A^T\mu, x^{k-1}-x^k\rangle = \langle \mu, A(x^{k-1}-x^k)\rangle = \langle \mu, b-b\rangle = 0$$

et nous obtenons donc la première égalité.

Alors, la combinaison de (3.18) avec (3.20) donne

$$f(x^k) \leq f(x^{k-1}) + \langle x^k - x^{k-1}, g^{k-1}\rangle \left(1 - \frac{L\lambda_k}{2\sigma}\right). \tag{3.22}$$

En effet, par (3.18), nous devons montrer que

$$\langle \nabla f(x^{k-1}), x^k - x^{k-1}\rangle + \frac{L}{2}\|x^k - x^{k-1}\|^2 \leq \langle x^k - x^{k-1}, g^{k-1}\rangle \left(1 - \frac{L\lambda_k}{2\sigma}\right).$$

Nous avons

$$\begin{aligned}&\langle \nabla f(x^{k-1}), x^k - x^{k-1}\rangle + \frac{L}{2}\|x^k - x^{k-1}\|^2 \\ &\leq \langle \nabla f(x^{k-1}), x^k - x^{k-1}\rangle + \frac{L}{2\sigma}\lambda_k\langle \nabla f(x^{k-1}), x^{k-1}-x^k\rangle \\ &\leq \langle \nabla f(x^{k-1}), x^{k-1}-x^k\rangle(-1) + \frac{L}{2\sigma}\lambda_k\langle \nabla f(x^{k-1}), x^{k-1}-x^k\rangle \\ &= -\langle g^{k-1}, x^{k-1}-x^k\rangle + \frac{L}{2\sigma}\lambda_k\langle g^{k-1}, x^{k-1}-x^k\rangle \\ &= \langle g^{k-1}, x^{k-1}-x^k\rangle\left(\frac{L\lambda_k}{2\sigma} - 1\right) \\ &= \langle x^{k-1}-x^k, g^{k-1}\rangle\left(\frac{L\lambda_k}{2\sigma} - 1\right) \\ &= \langle x^k - x^{k-1}, g^{k-1}\rangle\left(1 - \frac{L\lambda_k}{2\sigma}\right).\end{aligned}$$

Nous obtenons donc bien (3.22) de telle sorte que, avec $f_0(x^k) = f(x^k)$ et $f_0(x^{k-1}) = f(x^{k-1})$, nous avons

$$f_0(x^k) \leq f_0(x^{k-1}) + \langle x^k - x^{k-1}, g^{k-1} \rangle \left(1 - \frac{L\lambda_k}{2\sigma}\right).$$

Alors, avec $\lambda^* = \dfrac{2\varepsilon\sigma}{L}$, nous avons

$$f_0(x^k) \leq f_0(x^{k-1}) + \langle x^k - x^{k-1}, g^{k-1} \rangle (1 - \varepsilon)$$

car $\lambda_k \geq \lambda_*$ et $\lambda_* \in (0, \lambda^*)$ et donc $-\lambda_k \leq -\lambda^*$.

Ceci montre que (3.3) a lieu avec $m = 1 - \varepsilon$.
Donc, le résultat suit comme une conséquence directe du théorème 3.2.1. ▪

La seconde méthode étend celle proposée dans [1] et nous autorise à utiliser une règle de longueur de pas généralisée qui évoque celle utilisée dans la méthode du gradient classique, comme étudiée par Bertsekas [4].

Algorithme 2 : Règle de longueur de pas d'Armijo-Goldstein

Soient $\beta \in (0,1)$, $m \in (0,1)$ et $s > 0$ des scalaires choisis fixés.

Commencer à partir d'un point $x^0 \in C \cap \mathcal{V}$ et générer la suite $(x^k) \in C \cap \mathcal{V}$ comme suit :

- Si $\nabla f(x^{k-1}) \in \mathcal{V}_0^{\perp}$, STOP.

- Sinon, avec $x^k(\lambda) = u(\lambda \nabla f(x^{k-1}), x^{k-1})$, poser $\lambda_k = \beta^{j_k} s$, où j_k est le premier entier positif j tel que

$$f(x^k(\beta^j s)) - f(x^{k-1}) \leq m \langle \nabla f(x^{k-1}), x^k(\beta^j s) - x^{k-1} \rangle. \tag{3.23}$$

Alors, poser $x^k = x^k(\lambda_k)$.

Dans le but de montrer que cette règle de longueur de pas est bien définie, nous avons besoin de la proposition 3.3.1 et, pour prouver cette dernière, nous avons besoin du théorème de Lax-Milgram.

Théorème 3.3.2 ***Théorème de Lax-Milgram***

Soit $a(u,v)$ *une forme bilinéaire, continue et coercive.*

Alors, pour tout $\phi \in H'$ *, il existe* $u \in H$ *unique tel que*

$$a(u,v) = \langle \phi, v \rangle \qquad \forall v \in H.$$

De plus, si a *est symétrique, alors* u *est caractérisé par la propriété*

$$u \in H \quad et \quad \frac{1}{2}a(u,v) - \langle \phi, u \rangle = \min_{v \in H} \left\{ \frac{1}{2}a(v,v) - \langle \phi, v \rangle \right\}.$$

La preuve de ce théorème peut être trouvée dans [7]. ∎

Proposition 3.3.1 *Pour tout* $x \in C \cap \mathcal{V}$ *, tout* $v \in \mathbb{R}^n$ *et* $\lambda > 0$ *, l'unique solution* $u(\lambda v, x)$ *définie par (3.12) satisfait* $u(0,x) = x$ *et les propriétés suivantes ont lieu :*

1. $\sigma \|x - u(\lambda v, x)\|^2 \leq \lambda \langle x - u(\lambda v, x), v \rangle$,

2. $\dfrac{\|u(\lambda v, x) - x\|}{\lambda} \leq \dfrac{1}{\sigma}\|v\|$,

3. $\lim\limits_{\lambda \to 0+} \dfrac{u(\lambda v, x) - x}{\lambda}$ *existe et est égale à* $\rho(v,x) = u$,
   où $u \in \mathcal{V}_0$ *satisfait*
   $$Q(x)u + v \in \mathcal{V}_0^{\perp} \tag{3.24}$$
   avec $Q(x) = \nabla_1^2 d(x,x)$,

4. $\langle -\rho(v,x), v \rangle \geq \sigma \|\rho(v,x)\|^2$.

Preuve

• Pas 1 :

Fixons $x \in C \cap \mathcal{V}$.
Par (3.12), nous avons $u(0,x) = \arg\min\{d(z,x) \mid z \in \mathcal{V}\}$ et donc, par les conditions d'optimalité (3.13) avec $\mu = 0$, il suit que $u(0,x) = x$.
En effet, en prenant le lagrangien $L(z,\mu) := d(z,x) + Az - b$ et, en dérivant, nous obtenons $\nabla_1 d(u(0,x),x) = 0$ et le résultat suit.

De plus, par (3.13), nous avons

$$\langle \lambda v + \nabla_1 d(u(\lambda v,x),x), x - u(\lambda v,x)\rangle = 0$$

et par cette égalité, (1) suit immédiatement en utilisant l'inégalité de forte convexité (3.11) avec $y = x_1 = x$ et $x_2 = u(\lambda v,x)$.
En effet, nous obtenons

$$\langle \nabla_1 d(x,x) - \nabla_1 d(u(\lambda v,x),x), x - u(\lambda v,x)\rangle \geq \sigma \|x - u(\lambda v,x)\|^2$$

$$\Leftrightarrow \langle -\nabla_1 d(u(\lambda v,x),x), x - u(\lambda v,x)\rangle \geq \sigma \|x - u(\lambda v,x)\|^2.$$

Il nous reste donc à montrer que

$$\lambda \langle x - u(\lambda v,x), v\rangle = \langle -\nabla_1 d(u(\lambda v,x),x), x - u(\lambda v,x)\rangle.$$

Or,

$$\begin{aligned} &\langle \lambda v + \nabla_1 d(u(\lambda v,x),x), x - u(\lambda v,x)\rangle = 0 \\ &\quad \Leftrightarrow \quad \langle \lambda v, x - u(\lambda v,x)\rangle + \langle \nabla_1 d(u(\lambda v,x),x), x - u(\lambda v,x)\rangle = 0 \\ &\quad \Leftrightarrow \quad \langle \lambda v, x - u(\lambda v,x)\rangle = -\langle \nabla_1 d(u(\lambda v,x),x), x - u(\lambda v,x)\rangle \end{aligned}$$

et en utilisant différentes propriétés du produit scalaire, le résultat suit.

• Pas 2 :

La propriété (2) suit de (1) et de l'inégalité de Cauchy-Schwarz.

En effet,

$$\begin{aligned} \sigma\|x-u(\lambda v,x)\|^2 &\leq \lambda\langle x-u(\lambda v,x),v\rangle \\ &\leq \lambda|\langle x-u(\lambda v,x),v\rangle| \\ &\leq \lambda\|x-u(\lambda v,x)\|\|v\| \end{aligned}$$

$$\Leftrightarrow \qquad \frac{\|x-u(\lambda v,x)\|}{\lambda} \leq \frac{1}{\sigma}\|v\|.$$

• Pas 3 :

Comme $d(\cdot,y)$ est fortement convexe sur $C\cap\mathcal{V}$, il suit de (3.11) que

$$\langle Q(x)h,h\rangle \geq \sigma\|h\|^2 \qquad \forall h\in\mathcal{V}_0. \tag{3.25}$$

En effet, le résultat suit immédiatement de la proposition 1.3.2.

Comme conséquence du théorème de Lax-Milgram, (3.24) admet exactement une solution $\rho(v,x)$.
En effet, posons $a(u,v)=\langle Q(x)u,v\rangle$.
Nous avons alors $\forall z\in\mathcal{V}_0, \exists u\in\mathcal{V}_0$,

$$\begin{aligned} \langle Q(x)u+v,z\rangle=0 \quad &\Leftrightarrow \quad \langle Q(x)u,z\rangle=\langle -v,z\rangle \\ &\Leftrightarrow \quad a(u,v)=\langle -v,z\rangle \end{aligned}$$

et donc, u est solution unique de (3.24).

Notons que $\nabla_1 d(x,x)=0$. Alors, comme, par (3.13), nous avons

$$\lambda v+\nabla_1 d(u(\lambda v,x),x)+A^T\mu(\lambda v,x)=0, \tag{3.26}$$

il suit que

$$\forall h\in\mathcal{V}_0, \quad \frac{1}{\lambda}\langle\nabla_1 d(u(\lambda v,x),x)-\nabla_1 d(x,x),h\rangle=\langle -v,h\rangle.$$

En effet, ceci est équivalent à

$$\begin{aligned} &\frac{1}{\lambda}\langle\nabla_1 d(u(\lambda v,x),x),h\rangle=\langle -v,h\rangle \\ &\Leftrightarrow \qquad \langle\nabla_1 d(u(\lambda v,x),x),h\rangle=\lambda\langle -v,h\rangle=\lambda\langle -\lambda v,h\rangle. \end{aligned}$$

Or, par (3.26), nous avons

$$\begin{aligned}\langle \nabla_1 d(u(\lambda v, x), x), h\rangle &= \langle -\lambda v - A^T \mu(\lambda v, x), h\rangle \\ &= \langle -\lambda v, h\rangle - \langle A^T \mu(\lambda v, x), h\rangle\end{aligned}$$

et comme

$$\langle A^T \mu(\lambda v, x), h\rangle = \langle \mu(\lambda v, x), Ah\rangle = 0$$

car $h \in \mathcal{V}_0$, le résultat suit.

Notons $s(\lambda) := \dfrac{u(\lambda v, x)}{\lambda}$.

Maintenant, par (2), la suite généralisée $(s(\lambda))_{\lambda>0}$ est bornée.

Comme $u(\lambda v, x) = x + \lambda s(\lambda)$, en prenant la limite, en notant que $\lambda \to 0^+$ dans la dernière équation et en utilisant la définition de la dérivée (en se rappelant que $d(\cdot, y) \in C^2$ pour chaque $y \in C \cap \mathcal{V}$), il suit que toute valeur d'adhérence u de la suite généralisée $(s(\lambda))_{\lambda>0}$ satisfait $u \in \mathcal{V}_0$ tel que

$$\langle \nabla_1^2 d(x, x)u, h\rangle = \langle Q(x)u, h\rangle = -\langle v, h\rangle \quad \forall h \in \mathcal{V}_0,$$

ce qui est équivalent à (3.24).

Par conséquent, $u = \rho(v, x)$ et $\lim\limits_{\lambda \to 0^+} s(\lambda)$ existe et est égale à $\rho(v, x)$.

- Pas 4 :

Pour prouver (4), prendre $h = \rho(v, x) \in \mathcal{V}_0$ dans la dernière égalité et utiliser (3.25).
En effet, comme $h = \rho(v, x) = u$, nous avons

$$\langle Q(x)h, h\rangle = \langle -v, h\rangle = \langle h, -v\rangle = \langle -h, v\rangle$$

ce qui donne le résultat. ∎

Nous pouvons maintenant prouver la convergence de l'algorithme 2.

Théorème 3.3.3 *Soit* (x^k) *la suite générée par l'algorithme 2.*

Si, au pas k, *nous avons* $\nabla f(x^{k-1}) \in \mathcal{V}_0^{\perp}$, *alors* x^{k-1} *est une solution optimale.*

Sinon, l'algorithme est bien défini, i.e. il existe un entier j_k *tel que* $\lambda_k = \beta^{j_k}$ *et la suite* (λ_k) *est bornée par valeurs inférieures par*

$$\lambda_* = \min(2\sigma\beta L^{-1}(1-m), s) > 0.$$

De plus, le théorème 3.3.1 est valable pour la suite produite par l'algorithme 2.

Preuve

Nous avons juste à prouver que l'algorithme est bien défini et que $\lambda_k \geq \lambda_*$ (tel que $\lim_{n\to+\infty} \sigma_n = +\infty$). En effet, si nous posons $\varepsilon_k = 0$ et g^{k-1} comme donné dans (3.16), alors, par définition de l'algorithme 2, la suite (x^k) satisfait aux relations (3.1), (3.2) et (3.3).

Vérifions que ces trois relations sont bien satisfaites.

Comme (x^k) est générée par l'algorithme 2, $x^{k-1} \in C$ et nous avons $g^{k-1} \in \partial_0 f(x^{k-1})$ par (3.16) ce qui montre que la relation (3.1) est satisfaite.
La relation (3.2) suit en appliquant le même raisonnement que dans la preuve du théorème 3.3.1.
La relation (3.3) suit de l'algorithme 2 et de l'égalité (3.16).

Pour simplifier les notations, posons $x := x^{k-1}$, $v = \nabla f(x^{k-1})$ et $x(\lambda) = u(\lambda v, x)$.

Tout d'abord, si $v \in \mathcal{V}_0^{\perp}$, comme $x \in C \cap \mathcal{V}$, alors, par les conditions d'optimalité (3.13), il suit de l'algorithme que x est aussi une solution optimale.

Supposons maintenant que $v \notin \mathcal{V}_0^{\perp}$ et que (3.23) n'a pas lieu. Nous obtenons

$$f(x(\beta^j s)) - f(x) > m\langle x(\beta^j s) - x, v\rangle \qquad \forall j \in \mathbb{N}. \tag{3.27}$$

En invoquant le théorème des accroissements finis, $\exists\ z_j \in]x, x(\beta^j s)[$ tel que

$$\left\langle \nabla f(z_j), \frac{x(\beta^j s) - x}{\beta^j s} \right\rangle > m \left\langle \frac{x(\beta^j s) - x}{\beta^j s}, v \right\rangle \qquad \forall j \in \mathbb{N}. \tag{3.28}$$

En effet, $\exists\ z_j \in]x, x(\beta^j s)[$ tel que

$$f'(z_j) = \nabla f(z_j) = \frac{f(x(\beta^j s)) - f(x)}{x(\beta^j s) - x}.$$

D'où,

$$\begin{aligned} \langle \nabla f(z_j), x(\beta^j s) - x \rangle &= \left\langle \frac{f(x(\beta^j s)) - f(x)}{x(\beta^j s) - x}, x(\beta^j s) - x \right\rangle \\ &= f(x(\beta^j s)) - f(x) \\ &> m\langle x(\beta^j s) - x, v \rangle. \end{aligned}$$

Mais, par la proposition 3.3.1 (1), il suit que $\lim_{j\to\infty} z_j = x$.
En effet, nous savons que $z_j \in]x, x(\beta^j s)[$ avec $\beta \in (0,1), s > 0$ et j premier entier positif tel que (3.23) a lieu.
De plus, par l'algorithme 2, nous avons $x^k(\lambda) = u(\lambda \nabla f(x^{k-1}), x^{k-1})$ et $\lambda_k = \beta^{j_k} s$.
Donc, par la proposition et par l'inégalité de Cauchy-Schwarz, nous obtenons

$$\begin{aligned} \sigma \|x - x(\beta^j s)\|^2 &\leq \beta^j s \langle x - x(\beta^j s), \nabla f(x) \rangle \\ &\leq \beta^j s \|x - x(\beta^j s)\| \|\nabla f(x)\|. \end{aligned}$$

D'où, comme $\beta^j s \to 0$, par le théorème de l'étau,

$$\sigma \|x - x(\beta^j s)\| \to 0.$$

Ce qui donne finalement $x(\beta^j s) \to x$ et nous obtenons le résultat.

De plus, en passant à la limite dans (3.28) et en utilisant les points (3) et (4) de la même proposition, nous obtenons

$$\sigma(1-m)\|\rho(v,x)\|^2 \leq (1-m)\langle -v, \rho(v,x) \rangle \leq 0. \tag{3.29}$$

En effet, la première inégalité suit immédiatement de (4) et par propriété du produit scalaire.

Il reste donc à montrer que $\langle -v, \rho(v,x)\rangle \leq 0$ car $(1-m) > 0$.

Par (3), il suffit de montrer que

$$\begin{aligned}
\left\langle -v, \frac{u(\lambda v, x) - x}{\lambda} \right\rangle \leq 0 \qquad & (\lambda \to 0^+) \\
\Leftrightarrow \qquad \langle -\lambda v, u(\lambda v, x) - x\rangle \quad & \leq 0 \quad (\lambda \to 0^+) \\
\Leftrightarrow \qquad \langle \lambda v, x - u(\lambda v, x)\rangle \quad & \geq 0 \quad (\lambda \to 0^+) \\
\Leftrightarrow \qquad \lambda\langle v, x - u(\lambda v, x)\rangle \quad & \geq 0 \quad (\lambda \to 0^+)
\end{aligned}$$

ce qui est vrai par le point (1) de la proposition.

L'équation (3.29) implique que $\rho(v,x) = 0$ car $\sigma(1-m) > 0$ et donc, par (3.24), il suit que $v \in \mathcal{V}_0^\perp$ et nous avons atteint une contradiction.

Maintenant, prouvons que $\lambda_k \geq \lambda_*$.

Comme pour le cas de la règle de longueur de pas constant, en utilisant à nouveau le lemme de descente, nous obtenons

$$f_0(x^k(\lambda)) - f_0(x^{k-1}) \leq \langle g^{k-1}, x^k(\lambda) - x^{k-1}\rangle \left(1 - \frac{L\lambda}{2\sigma}\right), \qquad \forall \lambda > 0 \quad (3.30)$$

où $x^k(\lambda) = u(\lambda \nabla f(x^{k-1}), x^{k-1})$.
En effet, en prenant dans le lemme de descente $f = f_0$, $x + y = x^k(\lambda)$, $x = x^{k-1}$ et donc $y = x^k(\lambda) - x^{k-1}$, nous obtenons

$$f_0(x^k(\lambda)) - f_0(x^{k-1}) \leq \langle \nabla f(x^{k-1}), x^k(\lambda) - x^{k-1}\rangle + \frac{L}{2}\|x^k(\lambda) - x^{k-1}\|^2$$

où $x^k(\lambda) - x^{k-1} \in \mathcal{V}_0$.

Ensuite, par (3.14) et (3.16), nous trouvons que

$$\langle \nabla f(x^{k-1}), x^k(\lambda) - x^{k-1}\rangle = \langle g^{k-1}, x^k(\lambda) - x^{k-1}\rangle.$$

En remplaçant dans l'équation précédente, nous avons

$$f_0(x^k(\lambda)) - f_0(x^{k-1}) \leq \langle g^{k-1}, x^k(\lambda) - x^{k-1}\rangle + \frac{L}{2}\|x^k(\lambda) - x^{k-1}\|^2.$$

Et nous obtenons le résultat en utilisant l'équation (3.18)

$$\begin{aligned} f_0(x^k(\lambda)) - f_0(x^{k-1}) &\leq \langle g^{k-1}, x^k(\lambda) - x^{k-1}\rangle + \frac{L}{2\sigma}\lambda\langle g^{k-1}, x^{k-1} - x^k(\lambda)\rangle \\ &= \langle g^{k-1}, x^k(\lambda) - x^{k-1}\rangle + (-1)\frac{L}{2\sigma}\lambda\langle g^{k-1}, x^k(\lambda) - x^{k-1}\rangle \\ &= \langle g^{k-1}, x^k(\lambda) - x^{k-1}\rangle \left(1 - \frac{L\lambda}{2\sigma}\right). \end{aligned}$$

La relation (3.30) implique finalement que (3.23) a lieu pour tout $j \in \mathbb{N}$ avec $\beta^j s \leq 2\sigma L^{-1}(1-m)$.

En effet, comme nous avons $\lambda \equiv \beta^j s \leq \frac{2\sigma}{L}(1-m)$ ce qui est équivalent à

$$1 - \frac{L\lambda}{2\sigma} \geq 1 - 1 + m$$

et puisque le produit scalaire $\langle g^{k-1}, x^k(\lambda) - x^{k-1}\rangle$ est négatif, nous obtenons bien (3.23).

Mais, comme par définition, si $j_k \neq 0$, $\lambda_k\beta^{-1}$ ne satisfait pas (3.23), alors $\lambda_k\beta^{-1} > 2\sigma L^{-1}(1-m)$, il suit que $\lambda_k \geq \lambda_* \quad \forall k$.

A partir d'ici, nous pouvons procéder avec les mêmes affirmations et conclusions que dans le théorème 3.3.1 pour l'algorithme 2. ∎

Chapitre 4

Méthodes du gradient intérieur avec efficacité améliorée

4.1 Introduction

Dans ce chapitre, nous analyserons davantage le taux de convergence global des méthodes de gradient intérieur et nous proposerons un nouveau schéma intérieur qui améliore leur efficacité.

Nous savons que la méthode classique du gradient intérieur pour minimiser une fonction continûment différentiable sur $\mathbb{R}^n$ donne un taux de convergence global estimé $\mathcal{O}(k^{-1})$ pour les valeurs de la fonction.

Nesterov [19] a développé ce qu'il appelle « un algorithme optimal » pour la minimisation convexe et a été capable d'améliorer l'efficacité de la méthode du gradient en construisant une méthode aussi simple que cette dernière, mais avec un taux de convergence plus rapide $\mathcal{O}(k^{-2})$.
En s'inspirant de ce travail, il est donc naturel de se demander si ce genre de résultat peut être étendu aux méthodes du gradient intérieur et nous allons montrer que c'est bien le cas.

Nous proposons un algorithme qui fournit une extension naturelle aux résultats de Nesterov [19] et qui conduit à une méthode de gradient intérieur « optimale » pour les problèmes convexes coniques.

4.2 Etapes de la construction de l'algorithme

Considérons à nouveau le problème d'optimisation conique suivant :

$$\inf\{f(x) : x \in \overline{C} \cap \mathcal{V}\}, \qquad (M)$$

où $\mathcal{V} := \{x \in \mathbb{R}^n \mid Ax = b\}$ avec $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$, $n \geq m$ et $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe sci.

Supposons qu'il existe $x^0 \in$ dom $f \cap C : Ax^0 = b$.

Supposons aussi que f est continûment différentiable avec ∇f lipschitzien sur $C \cap \mathcal{V}$ et de constante de Lipschitz L, c'est-à-dire satisfaisant (3.10).

L'idée de base est de générer une suite de fonctions (q^k) qui approxime la fonction f de telle sorte que, à chaque pas $k \geq 0$, la différence $q^k(x) - f(x)$ est réduite d'une fraction $(1 - \alpha_k)$, où $\alpha_k \in [0, 1)$, qui est

$$q^{k+1}(x) - f(x) \leq (1 - \alpha_k)(q^k(x) - f(x)) \qquad \forall x \in \overline{C} \cap \mathcal{V}. \qquad (4.1)$$

Chaque fois que la propriété (4.1) a lieu, nous obtenons alors

$$q^k(x) - f(x) \leq \gamma_k(q^0(x) - f(x)) \qquad \forall x \in \overline{C} \cap \mathcal{V}, \qquad (4.2)$$

où

$$\gamma_k := \prod_{l=0}^{k-1}(1 - \alpha_l). \qquad (4.3)$$

Donc, si nous supposons que l'ensemble des solutions optimales X_* du problème (P) est non vide et si, au pas k, nous avons une suite $(x^k) \in C \cap \mathcal{V}$ telle que

$$f(x^k) \leq \inf_{z \in \overline{C} \cap \mathcal{V}} q^k(z) := q^k_*,$$

nous obtenons de (4.2) l'estimation du taux de convergence global

$$f(x^k) - f(x) \leq \gamma_k(q^0(x^*) - f(x^*)). \qquad (4.4)$$

A partir de la dernière inégalité, il suit que, si $\gamma_k \to 0$, alors, la suite (x^k) est une suite minimale pour f et le taux de convergence de $f(x^k)$ vers $f(x^*)$ est mesuré par la grandeur de γ_k.

Donc, pour construire des algorithmes basés sur le schéma ci-dessus proposé par Nesterov [19], nous avons besoin de

- générer une suite de fonctions appropriées $(q^k(\cdot))$;
- garantir que, à chaque itération k,

$$f(x^k) \leq \min_{z \in \overline{C} \cap \mathcal{V}} q^k(z) := q_*^k.$$

Commençons par la construction de la suite $(q^k(\cdot))$.

Dans ce but, nous prenons $d \equiv H \in \mathcal{D}(C)$, où H est une distance proximale de Bregman avec un noyau h tel que

(H_1) dom $h = \overline{C}$;

(H_2) h est σ-fortement convexe sur $C \cap \mathcal{V}$.

Pour chaque $k \geq 0$ et pour tout $x \in \overline{C} \cap \mathcal{V}$, nous construisons récursivement le suite $(q^k(x))$ via

$$q^0(x) = f(x^0) + cH(x, x^0); \tag{4.5}$$

$$q^{k+1}(x) = (1 - \alpha_k) q^k(x) + \alpha_k l_k(x, y^k); \tag{4.6}$$

$$l_k(x, y^k) = f(y^k) + \langle x - y^k, \nabla f(y^k) \rangle. \tag{4.7}$$

Ici, $c > 0$ et $\alpha_k \in [0, 1)$.

Le point x^0 est choisi tel que $x^0 \in C \cap \mathcal{V}$, tandis que le point $y^k \in C$ est arbitraire et sera généré d'une manière spécifique plus tard.

Nous montrons d'abord que la suite $(q^k(\cdot))$ satisfait (4.1).

Lemme 4.2.1 *La suite $(q^k(x))$ définie par (4.5), (4.6) et (4.7) satisfait*

$$q^{k+1}(x) - f(x) \leq (1 - \alpha_k)(q^k(x) - f(x)) \qquad \forall x \in \overline{C} \cap \mathcal{V}.$$

Preuve

Comme f est convexe, nous avons $f(x) \geq l_k(x, y^k)\ \forall x \in \overline{C} \cap \mathcal{V}$ ce qui, mis avec (4.6) nous donne

$$q^{k+1}(x) \leq (1 - \alpha_k) q^k(x) + \alpha_k f(x) \qquad \forall x \in \overline{C} \cap \mathcal{V},$$

et le résultat voulu suit.
En effet,

$$\begin{aligned} q^{k+1}(x) - f(x) &\leq (1 - \alpha_k) q^k(x) + \alpha_k f(x) - f(x) \\ &= (1 - \alpha_k) q^k(x) + (\alpha_k - 1) f(x) \\ &= (1 - \alpha_k)(q^k(x) + f(x)). \end{aligned}$$

■

En utilisant les notations du chapitre 3, nous rappelons que, pour chaque $z \in C \cap \mathcal{V}$ et pour chaque $v \in \mathbb{R}^n$, il existe un point $u(v, z) \in C \cap \mathcal{V}$ unique (par forte convexité de $H(\cdot, z)$) qui résout

$$u(v, z) = \arg\min\{\langle v, x\rangle + H(x, z) \mid x \in \overline{C} \cap \mathcal{V}\}. \tag{4.8}$$

Le prochain résultat est crucial et montre que la suite $(q^k(\cdot))$ admet une forme simple générique.

Lemme 4.2.2 *Pour tout $k \geq 0$, nous avons*

$$q^k(x) = q_*^k + c_k H(x, z^k) \qquad \forall x \in \overline{C} \cap \mathcal{V} \tag{4.9}$$

avec

$$z^k = \arg\min_{x \in \overline{C} \cap \mathcal{V}} q^k(x),\ q_*^k = q^k(z^k),\ c_0 = c,\ z^0 = x^0 \in C \cap \mathcal{V}. \tag{4.10}$$

De plus, la suite $(z^k) \in C \cap \mathcal{V}$ est uniquement définie par

$$\begin{aligned} z^{k+1} &= \arg\min \left\{ \left\langle x, \frac{\alpha_k}{c_{k+1}} \nabla f(y^k) \right\rangle + H(x, z^k) \mid x \in \overline{C} \cap \mathcal{V} \right\} \\ &\equiv u\left(\frac{\alpha_k}{c_{k+1}} \nabla f(y^k), z^k \right), \end{aligned} \tag{4.11}$$

où la suite positive (c_k) satisfait $c_{k+1} = (1 - \alpha_k) c_k$.

Preuve

La preuve se fait par récurrence et nous utiliserons l'identité des trois points (2.19) :

$$H(c,a) = H(c,b) + H(b,a) + \langle c-b, \nabla_1 H(b,a)\rangle \qquad \forall a,b \in C, \forall c \in \text{dom } f.$$

Pour $k = 0$, comme $z^0 = x^0$ par (4.5), nous avons $q^0(x) = f(x^0) + cH(x,z^0)$.

Alors, comme $c\nabla_1 H(z^0,z^0) = 0$, par les propriétés de H et comme $z^0 \in C \cap \mathcal{V}$, les conditions d'optimalité impliquent que

$$z^0 = \arg \min_{x \in \overline{C} \cap \mathcal{V}} q^0(x).$$

Supposons maintenant que (4.9) a lieu pour un certain k et montrons que, $\forall x \in \overline{C} \cap \mathcal{V}$,

$$q^{k+1}(x) = q_*^{k+1} + c_{k+1} H(x, z^{k+1}). \tag{4.12}$$

En substituant (4.9) dans (4.6) et en utilisant le fait que $c_{k+1} = (1-\alpha_k)c_k$, nous obtenons

$$q^{k+1}(x) = (1-\alpha_k) q_*^k + c_{k+1} H(x, z^k) + \alpha_k l_k(x, y^k). \tag{4.13}$$

Alors, par définition de z^{k+1}, nous avons

$$z^{k+1} = \arg \min_{x \in \overline{C} \cap \mathcal{V}} q^{k+1}(x) = u\left(\frac{\alpha_k}{c_{k+1}} \nabla f(y^k), z^k\right)$$

avec $z^{k+1} \in C \cap \mathcal{V}$ et

$$q_*^{k+1} = q^{k+1}(z^{k+1}) = (1-\alpha_k) q_*^k + c_{k+1} H(z^{k+1}, z^k) + \alpha_k l_k(z^{k+1}, y^k). \tag{4.14}$$

En soustrayant (4.14) de (4.13), i.e.

$$\begin{aligned} q^{k+1}(x) - q_*^{k+1} \;=\; & (1-\alpha_k) q_*^k + c_{k+1} H(x, z^k) + \alpha_k l_k(x, y^k) \\ & -(1-\alpha_k) q_*^k - c_{k+1} H(z^{k+1}, z^k) - \alpha_k l_k(z^{k+1}, y^k) \end{aligned}$$

$$\Leftrightarrow q^{k+1}(x) = q_*^{k+1} + c_{k+1}[H(x,z^k) - H(z^{k+1},z^k)] + \alpha_k [l_k(x,z^k) - l_k(z^{k+1},z^k)]$$

et en utilisant (4.7), nous obtenons

$q^{k+1}(x) =$

$$q_*^{k+1} + c_{k+1}[H(x,z^k) - H(z^{k+1},z^k)] + \alpha_k \langle z^{k+1} - x, -\nabla f(y^k)\rangle. \tag{4.15}$$

Maintenant, comme $z^{k+1} = \arg \min_{x \in \overline{C} \cap \mathcal{V}} q^{k+1}(x)$, alors en écrivant les conditions d'optimalité de (4.13) (et en se rappelant les propriétés de H), nous avons

$$c_{k+1}\langle \nabla_1 H(z^{k+1}, z^k), z^{k+1} - x\rangle = -\langle \alpha_k \nabla f(y^k), z^{k+1} - x\rangle, \tag{4.16}$$

$\forall x \in \overline{C} \cap \mathcal{V}$.

En utilisant (4.16) dans (4.15), il suit que, pour tout $x \in \overline{C} \cap \mathcal{V}$,
$q^{k+1}(x) =$

$$q_*^{k+1} + c_{k+1}[H(x, z^k) - H(z^{k+1}, z^k) + \langle z^{k+1} - x, \nabla_1 H(z^{k+1}, z^k)\rangle]. \tag{4.17}$$

En invoquant l'identité de départ avec $c = x$, $b = z^{k+1}$ et $a = z^k$, l'équation (4.17) se réduit à

$$q^{k+1}(x) = q_*^{k+1} + c_{k+1} H(x, z^{k+1})$$

et le lemme est prouvé. ∎

Le résultat suivant est fondamental pour déterminer les pas principaux de l'algorithme, à savoir, les formules utilisées pour mettre à jour la suite (x^k) et pour déterminer le choix du point intermédiaire y^k.

Théorème 4.2.1 *Soient $\sigma > 0$ et $L > 0$ donnés.*

Supposons que, pour un certain $k \geq 0$, nous avons un point $x^k \in C \cap \mathcal{V}$ tel que $f(x^k) \leq q_*^k = \min\{q^k(x) \mid x \in \overline{C} \cap \mathcal{V}\}$.

Soient $\alpha_k \in [0, 1), c_{k+1} = (1 - \alpha_k)c_k$ et une suite $(z^k) \in C \cap \mathcal{V}$ donnée par (4.11).

Définissons

$$y^k = (1 - \alpha_k)x^k + \alpha_k z^k; \tag{4.18}$$

$$x^{k+1} = (1 - \alpha_k)x^k + \alpha_k z^{k+1}. \tag{4.19}$$

Alors, les inégalités suivantes ont lieu :

$$q_*^{k+1} \geq f(x^{k+1}) \frac{1}{2} \left(\frac{c_{k+1}\sigma}{\alpha_k^2} - L \right) \|x^{k+1} - y^k\|^2.$$

Preuve

Soit $x^k \in C \cap \mathcal{V}$. Comme $q^k(x) = q_*^k + c_k H(x, z^k)$, alors, par (4.6) et en utilisant le fait que $c_{k+1} = (1 - \alpha_k)c_k$, nous avons

$$q^{k+1}(x) = (1 - \alpha_k)q_*^k + c_{k+1}H(x, z^k) + \alpha_k l_k(x, y^k)$$

et avec $z^{k+1} = \arg \min_{x^k \in C \cap \mathcal{V}} q^{k+1}(x)$, nous obtenons

$$q^{k+1}(z^{k+1}) = q_*^{k+1} = (1 - \alpha_k)q_*^k + c_{k+1}H(z^{k+1}, z^k) + \alpha_k l_k(z^{k+1}, y^k). \quad (4.20)$$

Sous nos hypothèses, nous avons $q_*^k \geq f(x^k)$ et donc, en utilisant l'inégalité du gradient pour f , nous avons

$$q_*^k \geq f(x^k) \geq f(y^k) + \langle x^k - y^k, \nabla f(y^k) \rangle.$$

Il suit de (4.7) et (4.20) que

$$q_*^{k+1} \geq f(y^k) + c_{k+1}H(z^{k+1}, z^k) + \langle \nabla f(y^k), r^k \rangle \quad (4.21)$$

où $r^k = \alpha_k(z^{k+1} - y^k) + (1 - \alpha_k)(x^k - y^k)$.
En effet,

$$\begin{aligned}
q_*^{k+1} &= (1 - \alpha_k)q_*^k + c_{k+1}H(z^{k+1}, z^k) + \alpha_k l_k(z^{k+1}, y^k) \\
&\geq (1 - \alpha_k)f(y^k) + (1 - \alpha_k)\langle x^k - y^k, \nabla f(y^k) \rangle + c_{k+1}H(z^{k+1}, z^k) \\
&\quad + \alpha_k l_k(z^{k+1}, y^k) \\
&= (1 - \alpha_k)f(y^k) + (1 - \alpha_k)\langle x^k - y^k, \nabla f(y^k) \rangle + c_{k+1}H(z^{k+1}, z^k) \\
&\quad + \alpha_k f(y^k) + \alpha_k \langle z^{k+1} - y^k, \nabla f(y^k) \rangle \\
&= f(y^k) + (1 - \alpha_k)\langle x^k - y^k, \nabla f(y^k) \rangle + c_{k+1}H(z^{k+1}, z^k) \\
&\quad + \alpha_k \langle z^{k+1} - y^k, \nabla f(y^k) \rangle \\
&= f(y^k) + c_{k+1}H(z^{k+1}, z^k) \\
&\quad + \langle \nabla f(y^k), (1 - \alpha_k)(x^k - y^k) + \alpha_k(z^{k+1} - y^k) \rangle \\
&= f(y^k) + c_{k+1}H(z^{k+1}, z^k) + \langle \nabla f(y^k), r^k \rangle.
\end{aligned}$$

Notons que r^k peut s'écrire

$$r^k = (1 - \alpha_k)x^k + \alpha_k z^k - y^k + \alpha_k(z^{k+1} - z^k).$$

En effet,

$$\begin{aligned} r^k &= (1-\alpha_k)(x^k - y^k) + \alpha_k(z^{k+1} - y^k) \\ &= (1-\alpha_k)\ x^k - y^k + \alpha_k y^k + \alpha_k z^{k+1} - \alpha_k y^k \\ &= (1-\alpha_k)\ x^k - y^k + \alpha_k z^k - \alpha_k z^k + \alpha_k z^{k+1} \\ &= (1-\alpha_k)\ x^k - y^k + \alpha_k z^k + \alpha_k(z^{k+1} - z^k). \end{aligned}$$

Ensuite, par définition, nous avons $(1-\alpha_k)x^k - y^k + \alpha_k z^k = 0$ et (4.21) peut être réduite à

$$q_*^{k+1} \geq f(y^k) + c_{k+1}H(z^{k+1}, z^k) + \langle \alpha_k(z^{k+1} - z^k), \nabla f(y^k)\rangle. \tag{4.22}$$

En utilisant les définitions de y^k et $x^{k+1} \in C \cap \mathcal{V}$ données en (4.18) et (4.19), nous avons $x^{k+1} - y^k = \alpha_k(z^{k+1} - z^k)$.

Comme, par (H_2), h est σ-fortement convexe, il suit que

$$H(z^{k+1}, z^k) \geq \frac{\sigma}{2}\|z^{k+1} - z^k\|^2$$

et à partir de (4.22), nous obtenons

$$q_*^{k+1} \geq f(y^k) + \frac{1}{2}\frac{c_{k+1}\sigma}{\alpha_k^2}\|x^{k+1} - y^k\|^2 + \langle \nabla f(y^k), x^{k+1} - y^k\rangle. \tag{4.23}$$

En effet,

$$\begin{aligned} q_*^{k+1} &\geq f(y^k) + c_{k+1}H(z^{k+1}, z^k) + \langle \alpha_k(z^{k+1} - z^k), \nabla f(y^k)\rangle \\ &= f(y^k) + c_{k+1}H(z^{k+1}, z^k) + \langle \nabla f(y^k), \alpha_k(z^{k+1} - z^k)\rangle \\ &= f(y^k) + c_{k+1}H(z^{k+1}, z^k) + \langle \nabla f(y^k), x^{k+1} - y^k\rangle \\ &\geq f(y^k) + c_{k+1}\frac{\sigma}{2}\|z^{k+1} - z^k\|^2 + \langle \nabla f(y^k), x^{k+1} - y^k\rangle \\ &= f(y^k) + c_{k+1}\frac{\sigma}{2}\left\|\frac{1}{\alpha_k}(x^{k+1} - y^k)\right\|^2 + \langle \nabla f(y^k), x^{k+1} - y^k\rangle \\ &= f(y^k) + \frac{1}{2}\frac{c_{k+1}\sigma}{\alpha_k^2}\|(x^{k+1} - y^k)\|^2 + \langle \nabla f(y^k), x^{k+1} - y^k\rangle. \end{aligned}$$

A présent, si nous supposons que f est dans $C^{1,1}(C \cap \mathcal{V})$ i.e. f est une fonction continûment différentiable dont le gradient est lipschitzien, alors, par le lemme de descente avec $y = x^{k+1}$ et $x = y^k$, nous avons

$$f(y^k) + \langle x^{k+1} - y^k, \nabla f(y^k)\rangle \geq f(x^{k+1}) - \frac{L}{2}\|x^{k+1} - y^k\|^2. \tag{4.24}$$

En combinant avec (4.23), nous obtenons

$$q_*^{k+1} \geq f(x^{k+1}) + \frac{1}{2}\left(\frac{c_{k+1}\sigma}{\alpha_k^2} - L\right)\|x^{k+1} - y^k\|^2. \qquad \blacksquare$$

Par déduction, en prenant la suite (α_k) avec $\sigma c_{k+1} \geq L\alpha_k^2$, nous pouvons garantir que $q_*^{k+1} \geq f(x^{k+1})$.

En particulier, nous pouvons choisir $L\alpha_k^2 = \sigma c_k(1-\alpha_k)$ et nous obtenons l'algorithme du gradient intérieur amélioré suivant.

4.3 Algorithme du gradient intérieur amélioré

Algorithme du gradient intérieur amélioré (IGA)

- Pas 0 :

* Choisir un point $x^0 \in C \cap \mathcal{V}$ et une constante $c > 0$.
* Définir $z^0 = x^0 = y^0, c_0 = c$ et $\lambda = \sigma L^{-1}$.

- Pas k :

Pour $k \geq 0$, calculer :

* $\alpha_k = \dfrac{\sqrt{(c_k\lambda)^2 + 4c_k\lambda} - \lambda c_k}{2}$;

* $y^k = (1-\alpha_k)x^k + \alpha_k z^k$;

* $c_{k+1} = (1-\alpha_k)c_k$;

* $z^{k+1} = \arg\min\limits_{x \in \overline{C} \cap \mathcal{V}} \left\{\left\langle x, \dfrac{\alpha_k}{c_{k+1}}\nabla f(y^k)\right\rangle + H(x, z^k)\right\}$

$= u\left(\dfrac{\alpha_k}{c_{k+1}}\nabla f(y^k), z^k\right)$;

* $x^{k+1} = (1-\alpha_k)x^k + \alpha_k z^{k+1}$.

Notons que le travail de calcul de cet algorithme est exactement le même que celui de la méthode du gradient intérieur dans la section précédente via le calcul de z^{k+1}.

Pour estimer le taux de convergence nous avons besoin du lemme suivant sur la suite (α_k).

Lemme 4.3.1 *Soient* $\lambda_k > 0$ *et* $c_k > 0$ *avec* $c_0 = c$.

Soit (α_k) *la suite définie par* $\alpha_k^2 = \lambda_k c_k(1-\alpha_k)$ *avec* $\alpha_k \in [0,1[$ *et* $c_{k+1} = (1-\alpha_k)c_k$.

Posons $\gamma_k := \prod_{l=0}^{k-1}(1-\alpha_l)$. *Alors,*

$$\gamma_k \leq \left(1 + \frac{\sqrt{c}}{2}\sum_{l=0}^{k-1}\sqrt{\lambda_l}\right)^{-2}.$$

En particulier, avec $\lambda_l = \lambda \ \forall l$, *nous avons* $\gamma_k \leq 4(k\sqrt{\lambda c}+2)^{-2}$.

Un preuve de ce lemme peut être trouvée dans [19]. ∎

Nous obtenons donc une méthode du gradient intérieur convergente avec un taux de convergence estimé amélioré.

Théorème 4.3.1 *Soient* (x^k) *et* (y^k) *les suites générées par l'algorithme IGA.*

Soit x^* *la solution optimale de* (P).

Alors, pour tout $k \geq 0$, *nous avons*

$$f(x^k) - f(x^*) \leq \frac{4L}{\sigma k^2 c} C(x^*, x^0) = \mathcal{O}\left(\frac{1}{k^2}\right)$$

où $C(x^*, x^0) = c_0 H(x^*, x^0) + f(x^0) - f(x^*)$.

De plus, la suite (x^k) *est minimisante, i.e.* $f(x^k) \to f(x^*)$.

Preuve

Par le lemme 4.2.1, la suite de fonctions $(q^k(\cdot))$ satisfait (4.1) et, par conséquent, (4.4) a lieu, i.e. en utilisant (4.5), nous avons

$$\begin{aligned} f(x^k) - f(x^*) &\leq \gamma_k(q^0(x^*) - f(x^*)) \\ &= \gamma_k(f(x^0) + c_0 H(x^*, x^0) - f(x^*)) \\ &= \gamma_k C(x^*, x^0). \end{aligned}$$

En prenant le lemme 4.3.1 avec $\lambda_k = \sigma L^{-1}$, nous obtenons

$$\begin{aligned} \gamma_k &\leq \frac{4}{\left(k\sqrt{\frac{\sigma}{L}}c + 2\right)^2} \\ &= \frac{4L}{(k\sqrt{\sigma c} + 2\sqrt{L})^2} \\ &= \frac{4L}{k^2\sigma c + 4L + 4k\sqrt{\sigma c L}} \\ &= \frac{4L}{k^2\sigma c} + 1 + \frac{L}{k\sqrt{\sigma c L}} \\ &\leq \frac{4L}{\sigma c k^2}. \end{aligned}$$

Ainsi, le résultat suit

$$f(x^k) - f(x^*) \leq \frac{4L}{\sigma k^2 c} C(x^*, x^0) = \mathcal{O}\left(\frac{1}{k^2}\right). \qquad \blacksquare$$

Dès lors, pour résoudre (P) à une précision $\varepsilon > 0$ près, nous n'avons pas besoin de plus de $[\mathcal{O}(\frac{1}{\sqrt{\varepsilon}})]$ itérations de l'algorithme IGA ce qui représente une diminution significative (d'un facteur racine carrée) en comparaison avec la méthode du gradient intérieur du chapitre précédent.

Conclusion

Dans ce mémoire, nous avons étudié deux schémas itératifs permettant de résoudre le problème de minimisation convexe :

$$f_* = \inf\{f(x) \mid x \in \overline{C}\}, \qquad (P)$$

où $\overline{C}$ représente la fermeture de C, un ensemble convexe ouvert non vide de $\mathbb{R}^n$ et où $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ est une fonction convexe, propre et semi-continue inférieurement.

Dans le cadre du premier schéma basé sur la distance proximale, nous avons généré une suite (x^k) via l'itération suivante

$$x^k \in \arg\min\{\lambda_k f(x) + d(x, x^{k-1}) \mid x \in \overline{C}\}, \quad k = 1, 2, \ldots \quad (\lambda_k > 0).$$

Nous avons ensuite constaté que le taux de convergence de cette méthode était de l'ordre $\mathcal{O}(\sigma_n^{-1})$ avec $\sigma_n = \sum_{k=1}^{n} \lambda_k, \quad \lambda_k > 0$.

Le second schéma était basé sur la méthode du sous-gradient et nous avons produit une suite (x^k) via

$$x^k \in \arg\min\{\lambda_k \langle g^{k-1}, x\rangle + d(x, x^{k-1}) \mid x \in \overline{C}\}, \quad k = 1, 2, \ldots$$

Dans ce cas, le taux de convergence était de l'ordre $\mathcal{O}(k^{-1})$.

En particulier, nous avons trouvé une classe d'algorithmes du gradient intérieur qui fournit un taux de convergence global estimé de l'ordre $\mathcal{O}(k^{-2})$.

Bibliographie

[1] A. AUSLENDER and M. TEBOULLE. Interior gradient and epsilon-subgradient methods for constrained convex minimization. *Math. Oper. Res.*, 29 : 1–26, 2004.

[2] A. AUSLENDER and M. TEBOULLE. Interior gradient and proximal methods for convex and conic optimization. *SIAM J. Optim.*, 16(3) : 697–725, 2006.

[3] A. AUSLENDER, M. TEBOULLE, and S. BEN-TIBA. Interior proximal and multiplier methods based on second order homogeneous kernels. *Math. Oper. Res.*, 24 : 645–668, 1999.

[4] D.P. BERTSEKAS. On the Goldstein-Levitin-Polyak gradient projection method. *IEEE Trans. Automat. Control*, 21 : 174–183, 1976.

[5] D.P. BERTSEKAS. *Nonlinear Programming - 2d edition.* Athena Scientific, Belmont, MA, 1999.

[6] J. BOLTE and M. TEBOULLE. Barrier operators and associated gradient-like dynamical systems for constrained minimization problems. *SIAM J. Control Optim.*, 42 : 1266–1292, 2003.

[7] H. BREZIS. *Analyse Fonctionnelle : Théorie et Applications.* Masson, Paris, 1987.

[8] F. CALLIER. *Topologie - Notes de cours.* Département de mathématiques - FUNDP Namur, 2003-2004.

[9] Y. CENSOR and S. ZENIOS. The proximal minimization algorithm with d-functions. *J. Optim. Theory Appl.*, 73 : 451–464, 1992.

[10] G. CHEN and M. TEBOULLE. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim.*, 3 : 538–543, 1993.

[11] R. CORREA and C. LEMARECHAL. Convergence of some algorithm for convex programming. *Math. Program.*, 62 : 261–275, 1993.

[12] M. DOLJANSKY and M. TEBOULLE. An interior proximal algorithm and the exponential multiplier method for semidefinite programming. *SIAM J. Optim.*, 9 : 1–13, 1998.

[13] J. ECKSTEIN. Approximate iterations in Bregman function-based proximal algorithms. *Math. Program.*, 62 : 113–123, 1998.

[14] O. GÜLER. On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim.*, 29 : 403–419, 1991.

[15] J.-B. HIRIART-URRUTY and Cl. LEMARECHAL. *Convex Analysis and Minimization Algorithms I.* Springer-Verlag, New-York, 1993.

[16] K.C. KIWIEL. Proximal minimization methods with generalized Bregman functions. *SIAM J. Control Optim.*, 35 : 1142–1168, 1997.

[17] B. LEMAIRE. The proximal algorithm. *in New Methods in Optimization and Their Industrial Uses, Internat. Schriftenreihe Numer. Math. 87, J.P. Penot, ed., Birkhaüser, Basel*, pages 73–87, 1989.

[18] B. MARTINET. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle*, 4 : 154–158, 1970.

[19] Y. NESTEROV. On the approach to the construction of optimal methods of minimization of smooth convex functions. *Ekonom. i Mat. Metody*, 24 : 509–517, 1998.

[20] Y. NESTEROV and A. NEMIROVSKII. Interior point polynomial algorithms in convex programming. *SIAM*, 1994.

[21] J. RENEGAR. *A Mathematical View of Interior Methods in Convex Optimization.* MPS-SIAM Series of Optimization, Cornell University, Ithaca, New Jersey, 2001.

[22] S.M. ROBINSON. Linear convergence of epsilon subgradients methods for a class of convex functions. *Math. Program.*, 86 : 41–50, 1999.

[23] R. T. ROCKAFELLAR. *Convex Analysis.* Princeton University Press, Princeton, New Jersey, 1970.

[24] J.-J. STRODIOT. *Inexact Proximal Point Methods.* CIUF-CUD Summer School on Optimization and Applied Mathematics, Nha Trang University, Vietnam, 1-15 août 2004.

[25] J.-J. STRODIOT. *An Introduction to Nonsmooth Optimization.* CIUF-CUD Summer School on Optimization and Applied Mathematics, Can Tho University, Vietnam, 2-18 août 2003.

[26] J.-J. STRODIOT. *An Introduction to Optimization - Cours de première licence.* Département de Mathématiques - FUNDP Namur, 2005 - 2006.

[27] J.-J. STRODIOT. *Numerical Methods in Optimization - Cours de seconde licence.* Département de Mathématiques - FUNDP Namur, 2006-2007.

[28] J.-J. STRODIOT. *Interior-Point Methods.* CIUF-CUD Summer School on Optimization and Applied Mathematics, Nha Trang University, Vietnam, août 2002.

[29] M. TEBOULLE. Convergence of proximal-like algorithms. *SIAM J. Optim.*, 7 : 1069–1083, 1997.